UNIVERSITY OF LONDON THESIS

Degree PhD Year 2008 Name of Author CARDIN, Velia Maria De La Paz

THE FORM PATHWAYS

IN THE

VISUAL BRAIN

Velia María de la Paz

Cardin

UnIversity College London

PhD

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

I, Velia María de la Paz Cardin, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

The perception of visual forms is crucial for humans for successful interactions with the environment. This process occurs automatically, and its outcome is reflected in the inferences and decisions we constantly make. The focus of this thesis is on how the brain handles different aspects of the perception of forms. To study this in normal human individuals, experiments were performed using functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) and psychophysical methods. This thesis first discusses experiments designed to unravel the mechanisms of form construction, i.e. those from which all the component parts of a single form are assembled. Results suggest that the construction of very simple forms occurs in intermediate visual areas in a parallel and recursive process, with an increase in brain activity with increments in form complexity. A further experiment was performed to study how regularities or known characteristics of images, and the brain responses they elicit, will contribute to explain current percepts. Results from this experiment are consistent with a model where images with learnt attributes activate more strongly anterior visual areas and images with random patterns cause higher activations in earlier visual areas, probably due to top-down signals that reduce activity when it is possible to explain the causes of the sensory stimulation. Finally, it shows differences in the evoked neural activity when forms are either detected or classified, relating these processes to the activity generated in early visual areas. Based on the results of these experiments, a mechanism of top-down and bottom-up interactions between visual areas in the human brain is discussed in the context of the perception of forms.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

"An individual's life consisted of certain classified things: "real things" which were unfrequent and priceless, simply "things" which formed the routine stuff of life; and "ghost things," also called "fogs," such as fever, toothache, dreadful disappointments, and death. Three or more things occurring at the same time formed a "tower," or, if they came in immediate succession, they made a "bridge." "Real towers" and "real bridges" were the joys of life, and when the towers came in a series, one experienced supreme rapture; it almost never happened, though. In some circumstances, in a certain light, a neutral "thing" might look or even actually become "real" or else, conversely, it might coagulate into a fetid "fog." When the joy and the joyless happened to be intermixed, simultaneously or along the ramp of duration, one was confronted with "ruined towers" and "broken bridges."

Vladimir Nabokov. *Ada or ardor.*

This thesis is the result of the work done during my PhD, which was often difficult and challenging, full of "ruined towers" and "broken bridges", but had several "real things" as well. Its completion would not have been possible without several people who supported and helped me through these years.

Firstly, I would like to mention Professor Semir Zeki, who supervised my PhD, gave me all the freedom to pursue my research interests and always challenged my views. Working in his laboratory has been a wonderful learning experience, the results of which I am very pleased with.

I really appreciate all the advice from my secondary supervisor, Professor Karl Friston, who showed me in practice that implicit learning works, and also contributed to a huge amount of explicit learning as well. I really enjoyed attending his Monday meetings, where I saw many people interested in finding out how the brain works and making science in such an enthusiastic and ethical way.

I would also like to thank the Wellcome Trust, for generously funding my PhD, and Professor David Attwell, the director of the PhD programme, who mentored me (and many other people!) through all these years.

All the people in the Laboratory of Neurobiology, who made it such a nice and fun place to work, and were such an incredible help, from the discussion of experiments to making the printer work, including trips to Copenhagen, lab parties, pubs crawling, and obsessive Google searches about some fanatic characters of contemporary history. This would have been much more difficult, maybe even not possible, without you.

Finally, I would like to thank all the people that constitute the "real things" in my life.

I would like to start with my friends in London, with whom I share many of the uncertainties and rewards of the life in this city, who could somehow see through that moody face I sometimes have and embrace me so warmly from the very beginning, to have such a great time together.

I really want to thank my housemates in Ferndene Road and Gateway Mews, who always made me feel "at home", preparing delicious warm suppers which I very much appreciated after arriving from long days at work and cycle rides through the city. They tediously corrected my written and spoken English, and introduced me to great music, literature, cultures and quite different ways of seeing the world.

My friends elsewhere in the world, whom I constantly miss and wish they were here, so I could take them to see all the space invaders around the city, but mainly share very long conversations face to face, and not only over the phone.

My family, also disperse all around the planet, for being always extremely supportive and caring. I would specially like to thank my siblings, Mariano and María Victoria, who I very much admire for their strength, determination and big hearts, and my parents, Gracia Martinelli and Raúl Cardin, who raised me as an independent and critical person, and whose only motivation it has always been seeing me happy. I cannot really express how much I treasure them and all their love.

Part 1– INTRODUCTION

Chapter 1.1 – DESCRIPTION OF THE PROBLEM AND OVERVIEW

"My pockets always contain a rasping sediment of sand because I fill them with shells when I go on to the beach. The vast majority of these shells are round, sculptural forms the colour of a brown egg, with warm, creamy insides. They have a classical simplicity. The scarcely perceptible indentations of their surfaces flow together to produce a texture as subtly matt as that of a petal which is as satisfying to touch as Japanese skin. But there are also pure white shells heavily ridged on the outside but within of a marmoreal smoothness and these come in hinged pairs.
There is still a third kind of shell, though I find these less often. They are curlicued, shaped like turbans and dappled with pink, of a substance so thin the ocean easily grinds ways the outer husk to lay bare their spiralline cores. They are often decorated with baroque, infinitesimal swags of calcified parasites. They are the smallest of all the shells but by far the most intricate.
When I picked up one of those shells, I found it contained the bright pink, dried, detached limb of a tiny sea creature like a dehydrated memory. Sometimes a litter of dropped fish lies among the shells. Each fish reflects the sky with the absolute purity of a Taoist mirror."

Angela Carter. The smile of winter

Our visual experience of the world is full of forms, colours, textures, sizes, movements and concepts, which we perceive harmoniously, usually without confounding different parts with each other or with the background. From the sensory information we acquire through our eyes, we construct forms, we group some of them together into bigger wholes, we segregate each of these from the background and we associate them to learnt semantic concepts. If we analyse each individual entity, we can divide it into parts, and those into subparts, and even within the subparts we can perceive different textures and small edges. In the extract from Angela Carter's *The smile of winter*, we can see very clearly how the author perceives colours (i.e. brown, pink), shapes (i.e. spiralline, round) and textures (i.e. indented, marmoreal), groups them into bigger wholes (i.e. shell, fish), distinguishes the objects from the background (i.e. sky, sand) and associates them with concepts (i.e. sky, sand, shell, fish). We do all this automatically all the time; therefore it could seem an easy task. However, the problems the visual

brain has to solve when analysing the incoming information are computationally highly difficult. For example, when we look at the image in Fig 1.1.1, we can tell fairly easy that there are four flamingos in a lake. Making computational models that can arrive at the same conclusion has been extremely challenging for the computational neuroscience world - all these processes involved in the perception of form are not at all trivial for a computer, whereas the brain solves them with outstanding accuracy.

Fig. 1.1.1. Some of the problems the visual system outstandingly solves. When we look at this image, we know automatically it shows four flamingos and their reflections in the lake. By seeing the reflection in the background, we can even infer there is another flamingo. A computer program will have problems distinguishing the flamingos from their reflections or grouping the legs and the rest of the body into a whole object. Knowing that the reflection in the background could be caused by a flamingo, is far beyond the scope of current technology. Photograph by Anup and Manoj Shah, National Geographic Society ©.

We can start with a very simple three-stage model in trying to understand how form is processed in the brain (Fig 1.1.2): arrival of sensory information, form construction and recognition. The aim of this thesis is to study the neural mechanisms behind these processes. "Form" is a word used in many disciplines, making its definition difficult (see Uttal, 1988). This is not only the case for experts from diverse backgrounds such as art and science, it is also a debate within cognitive neuroscientists working on that specific aspect of visual perception. The usual approach has been to define form operationally. In this thesis, 'form' refers to "any segregated whole or unit", as defined by the Gestalt school (Kohler, 1929).

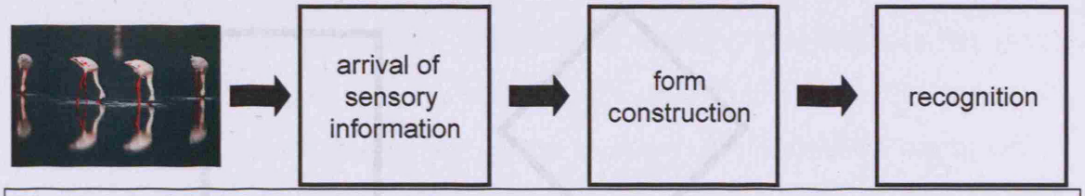


Fig 1.1.2. Simple model of form processing in the visual brain. In this simple model, visual information arrives at the cortex, the structural information about form is extracted and then associated to learnt concepts in a recognition stage. This model shows form processing as a straightforward mechanism, where information flows from one stage to the other. The work presented in this thesis actually shows that this is a much more complicated system, where information flows in recurrent loops within and between stages, for the system to determine the causes of the sensory stimulation in a more efficient way.

The Gestalt definition was chosen because it is the one that best explains the kind of issues and problems that will be discussed in this thesis. We can say that any spatially continuous independent visual element is a "form", but many of those forms, as we will see in further chapters, are grouped together to make larger ones, which are often not spatially continuous (see for example the stimuli in the second experiment). Therefore, "form" refers to every individual element as well as to the bigger wholes, as defined above. Related but different concepts are "figure" and "shape". Figure can have many definitions, but in the visual neurosciences it is usually referred to as

that part of the scene that is segregated from the background, generally referring to the whole, but not to the smaller components of objects or figures which are also forms. I do not use the term shape either, since it is usually defined as the spatial relationships of an object's elements. For example, Palmer (1999, p.p. 367) defines it as "the spatial structure of an object that does not change when some spatial transformations are applied to it". He says that if we have a square and a diamond (Fig 1.1.3), the same form is perceived as having two different shapes. This reflects the importance of the spatial location for the concept of shape, and probably the neural mechanisms behind it.

Fig 1.1.3. Same form, different shape. The same form is perceived as a square when its sides are vertical and horizontal, but as a diamond when it is rotated 45 degrees. Modified from Palmer, 1999, pp 367.

The Historical Survey, in the first part of this thesis, is a review of the literature of form perception from the psychological and neuroscientific points of view. I start with an account of functional specialization in the brain, emphasizing the architecture of the visual system and different levels of segregation of the sensory information in the visual pathway, including that related to form. There follows a description of the types of backward and forward connections that would mediate top-down and bottom-up mechanisms, which are thoroughly discussed in the Results and General Discussion sections, and it concludes with an account of the history of functional specialization and how this evolved into the modern view we have of the brain. The second part of the historical survey is a review of the experimental evidence of form and object perception from neuropsychology, experimental psychology, monkey electrophysiology and brain imaging.

With that background, the following section describes the result of the study of several aspects of the form pathway in humans.

The Methodology and Results part describes and discusses three experiments addressing different aspects of form perception. The perception of a form as a single, independent unit, requires the assembly of all its individual components (i.e. all the oriented lines that form a square or all the pixels that make an oriented line on a computer screen), a process I will refer to as form construction and which is the focus of Chapter 2.1. The next chapter (2.2) describes the brain areas activated by forms with different levels of extractable patterns, and how they interact through bottom-up and top-down mechanisms. Chapter 2.3 covers the neural mechanisms that support different levels of form recognition: a) detection, which is the ability to recognize there is something in the scene; and b) classification, the ability to put it into a conceptual group. Each chapter of this section starts with a brief introduction to the problem, a description of the methodology specific for the experiment, an account of the results and conclusions. A description of the methodology used in all the experiments can be found in Chapter 2.1. General information about some of the techniques and analysis methods can be found in the appendices.

Finally, the General Conclusion integrates the results of all the experiments in a theory of form perception in the visual brain, mediated by several visual areas interacting through top-down and bottom-up mechanisms.

Chapter 1.2. – HISTORICAL SURVEY

1.2.1. – FUNCTIONAL SPECIALIZATION IN THE VISUAL BRAIN

ANATOMY OF THE FORM SYSTEM

The work in this thesis deals with the visual pathways of form perception, whose function, as one of any sensory process in the brain, is to interpret the information that arrives to the cortex from the outside world. In this section I will review the anatomy of the visual system, and how this anatomical arrangement results in the segregation and parallel processing of visual information, including that of form.

Visual information reaches the brain through the optic nerves, that transmit the information from the retinal ganglion cells to the cortex via two main pathways:

a) the *tectal pathway*, where information goes from the retina to the superior colliculus, and from here to all the visual areas via the pulvinar or the lateral geniculate nucleus (LGN).

b) the *geniculate pathway*, in which 90% of the axons sent by ganglion cells go to LGN of the thalamus and then to the primary visual cortex in the occipital lobe. Some of them, however, reach the prestriate cortex directly (Yukie and Iwai, 1981).

Through these pathways, information is distributed to multiple visual centres in the brain, including the whole occipital lobe, and several parietal, temporal and frontal regions.

Visual information is transmitted in a segregated manner all the way from the retina to the cortex. Each retina receives information from the right and left part of the visual space. Information from each of the hemifields converges in the optic chiasm. This is where those fibres coming from the retina of the right eye that receive information from the right visual hemifield only, and those from the left retina that receive information from the left visual hemifield only, cross each other and join the fibres of the opposite eye that contain information about the same region in space (Fig 1.2.1). They

compose what it is known as the optic tract, which only carries inputs from the contralateral side of visual scene.

Projections from the LGN reach a region of the occipital cortex of the same hemisphere known as the primary visual cortex (V1), located on both sides of the calcarine sulcus. Since the LGN processes information of the contralateral hemifield, the same happens in the primary visual cortex. Topography is preserved through all these stages, where adjacent points in the retina are mapped into adjacent points in the LGN and adjacent points in V1 (and further visual areas!). The topographical mapping preserves qualitative spatial relations but not quantitative ones, with the central part of the visual field having proportionally a much greater representation than the periphery, a characteristic known as *cortical magnification factor* (Poliak, 1932; Talbot and Marshall, 1941; Daniel and Whitteridge, 1961). In each hemisphere, the upper visual hemifield is represented in the ventral lip of the calcarine sulcus, while the lower visual hemifield is represented in its dorsal lip. Central representations of the visual scene (5 degrees), corresponding to the fovea in the retina, are mapped into the occipital pole. The primary visual cortex it is also known as the striate cortex, since it has a very distinguishable cytoarchitecture of prominent stripes of white matter in layer 4, known as *Stria of Gennari*. The primary visual cortex projects to other topographically organized visual areas in the occipital lobe, known as the extrastriate areas.

An independent topographic map is not the only attribute that distinguishes these areas from each other. There are also distinct cytoarchitectonic patterns revealed by cytochrome oxidase staining (Horton, 1984; Livingstone and Hubel, 1988), which reveals the histological consequence of the segregated processing of information, preserved from the retina. Ganglion cells in the retina are subdivided into M and P cells, which transmit signals about motion and colour, respectively (Fig 1.2.2). The majority of the fibres of ganglion cells will relay in the LGN, where information is still segregated into colour and motion in the magnocellular and parvocellular pathway, and also into ocular specific layers. Segregation is also present in the cortex: in V1 information about colour and form is processed in blobs and interblobs of supragranular layers, respectively, whereas motion

information is processed in layer 4B. In V2, cytochrome oxidase staining reveals dark thick and thin stripes and paler interstripes which process motion, colour and form information, respectively, and project to different visual areas (Livingstone and Hubel, 1988; Shipp and Zeki, 1995).

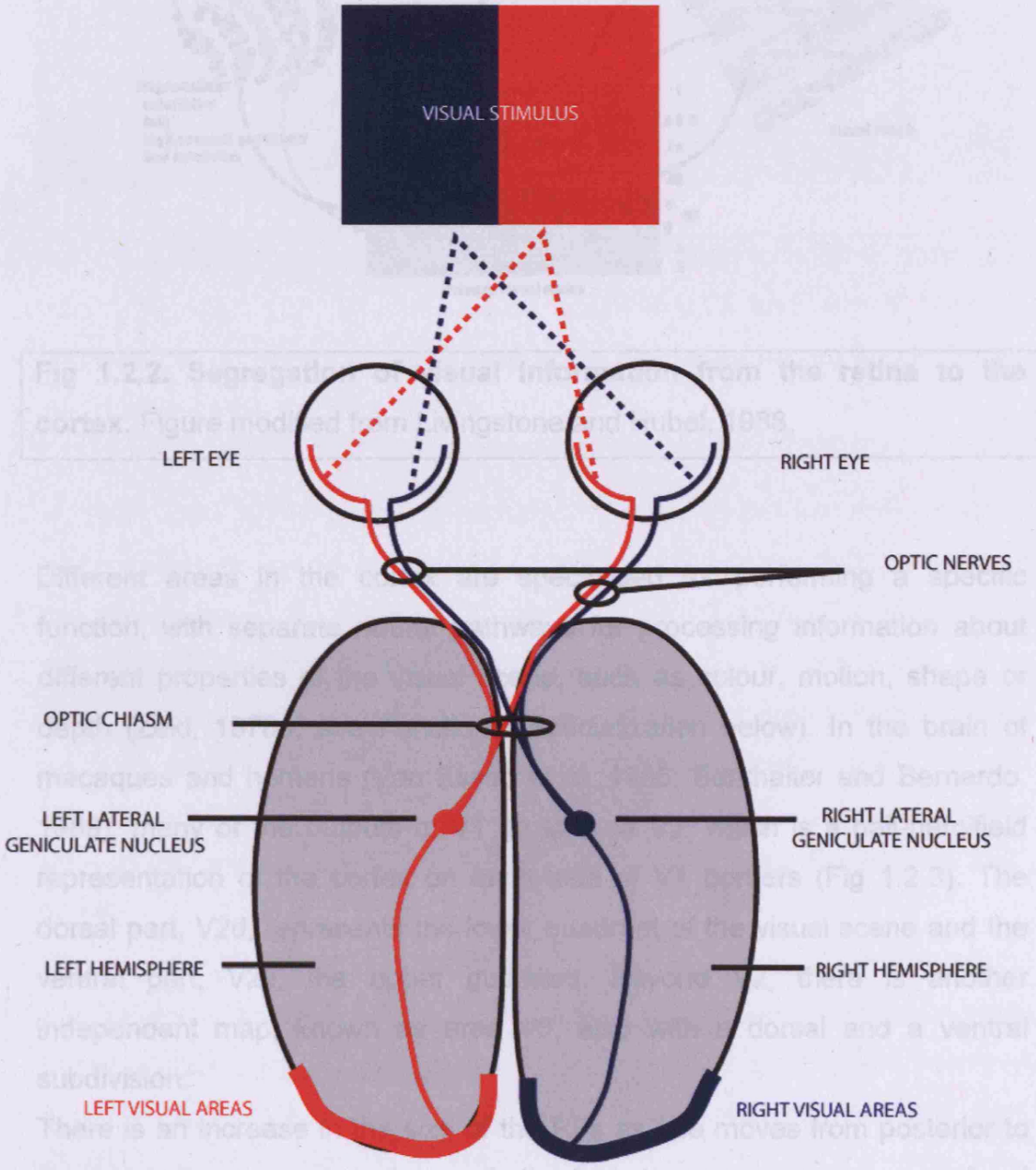


Fig 1.2.1. Anatomy of the visual system.

Fig 1.2.2. Segregation of visual information from the retina to the cortex. Figure modified from Livingstone and Hubel, 1988.

Different areas in the cortex are specialised for performing a specific function, with separate neural pathways for processing information about different properties of the visual scene, such as colour, motion, shape or depth (Zeki, 1978a; see Functional Specialization below). In the brain of macaques and humans (Van Essen et al, 1986; Burkhalter and Bernardo, 1989), many of the outputs of V1 go to area V2, which is a half-hemifield representation of the cortex on each side of V1 borders (Fig 1.2.3). The dorsal part, V2d, represents the lower quadrant of the visual scene and the ventral part, V2v, the upper quadrant. Beyond V2, there is another independent map, known as area V3, also with a dorsal and a ventral subdivision.

There is an increase in the size of the RFs as one moves from posterior to more anterior areas, and whereas in the first stages responses are confined to the contralateral visual hemifield, neurons at higher levels in temporal and parietal areas have receptive fields (RFs) of 10 by 10 degrees that usually include the fovea and extend over both halves of the visual field (Gross et

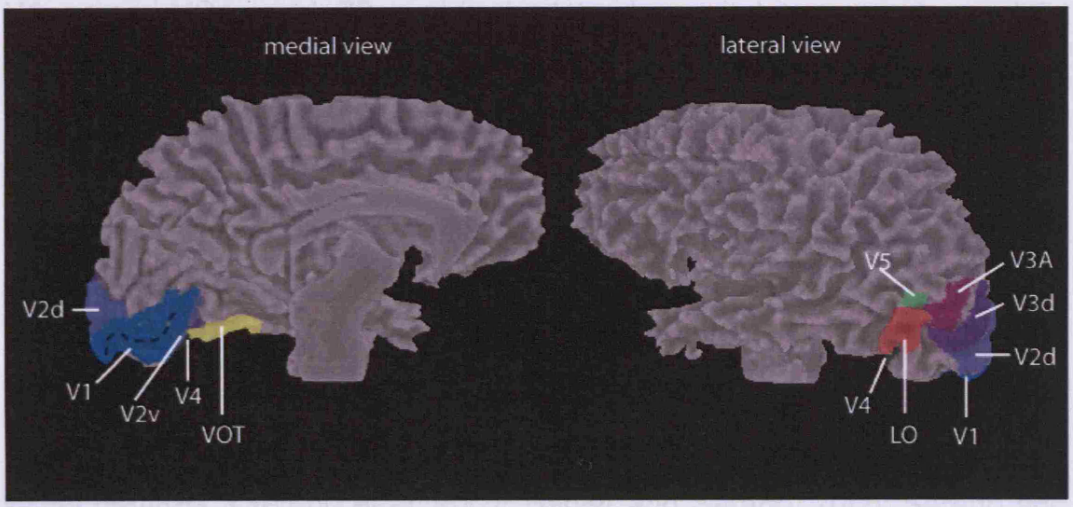


Fig 1.2.3. Visual areas in the human brain. Rough location of areas V1, V2d, V2v, V3v, V3d, V4, V5, lateral occipital (LO) and ventral occipito-temporal (VOT) cortex in the human brain. The blue dashed-line indicates the location of the calcarine sulcus.

al, 1969; Zeki 1974). It can be deduced from the above that differences in cytoarchitecture and an independent representation of the retina are two of the criteria taken into account to define a visual area. The inputs and outputs pattern and a specialization of function are other criteria that define them (see the section Functional specialization in the visual brain). Based on these criteria, several other areas have been described, like V3A, V4 and V5 (Fig. 1.2.3), which are associated with form, colour and motion processing, respectively (Zeki, 1974; Zeki, 1978b; Van Essen and Zeki, 1978).

With the incorporation of brain imaging techniques, it was possible to extend the findings in the brain of macaque monkeys to the mapping of visual areas in the human brain. Engel et al (1994) introduced the use of periodic stimuli to map the eccentricity and polar angle components of the visual field, allowing to functionally map visual areas in the human occipital lobe, including V1, V2, V3, V3A and V4, analogous to those in the macaque brain (Fig 1.2.3) (Sereno et al, 1995; DeYoe et al, 1996; Engel et al, 1997). More detailed mapping, using a variety of visual stimuli, allowed the description of further retinotopic areas in the ventral occipito-temporal cortex anterior to

V4, named VO1 and VO2, and in the lateral occipital cortex, between V3d and V5, named LO1 and LO2 (Brewer et al, 2005; Larsson and Heeger, 2006). Each of these areas contains a topographic representation of the contralateral hemifield and they are all mainly responsive to object-containing stimuli, probably processing shape information. Topographic organization in response to visual stimuli have also been shown in parietal and frontal areas with cyclic stimuli and tasks including attention, working-memory and spatial integration, showing that there is a topographic visual organization in higher areas for the further processing and integration of visual features (Sereno et al, 2001; Hagler and Sereno, 2006; Sereno and Huang, 2006).

TYPE OF CONNECTIONS AND THEIR ROLE

Everywhere in the cortex, there is constancy in the histological organization and in the connections between and within areas. This organization is the anatomical basis of top-down, bottom-up and local processing, which will be thoroughly discussed in this thesis, since they mediate all sensory and cognitive functions, and many of the conclusions of this work relate to the recurrent loops that arise through these connections.

Neurons in the brain can be divided, broadly speaking, into excitatory or inhibitory depending on whether they release glutamate or GABA, respectively, although there are neurons that release many other neurotransmitters with different modulatory functions.

In the visual cortex we find three main populations of neurons:

- Pyramidal cells - which are the main excitatory neurons in the cortex, extending through several layers of the cortex (see below) and with axons that leave the cortex towards other regions of the brain.
- Interneurons - local GABAergic cells of different shapes and locations, coupled through electric synapses.
- Spiny stellate cells - excitatory neurons with axons that terminate locally.

These types of neurons are distributed in a defined way in the cortex, which results in a histologically distinguishable laminar structure (Hubel and

Wiesel, 1972; Lund, 1988). In the visual cortex, six defined layers can be found, which contain different cellular types and projections:

- The *molecular layer 1*, contains very few neurons but a dense network of synapses, being therefore the location for many cortical computations that will have an effect on the firing properties of cells in deeper layers.
- The *supragranular layers (2, 3A and 3B)* contain many somata and dendrites of pyramidal cells; they receive little thalamic inputs and strong input from deeper areas. They contain many axons and dendrites of neurons located in all other cortical layers.
- *Layer 4*, known as the *granular layer*, it is mainly composed of pyramidal and stellate neurons and, in the primary visual cortex, it is further divided into four sublayers: 4A, 4B, 4Cα and 4Cβ. Layers 4Cα and 4Cβ are the main target of thalamo-cortical afferents as well as intra-hemispheric cortico-cortical afferents.
- *Layer 5, internal pyramidal*, contains the soma of large pyramidal cells and has dendrites extending through other layers and axons that will leave the cortex to other areas.
- Layer 6, multiform layer, has an heterogeneous population of neurons, and blends with the white matter by sending projections to the thalamus and other cortical areas, closing the loops of transmission of information.

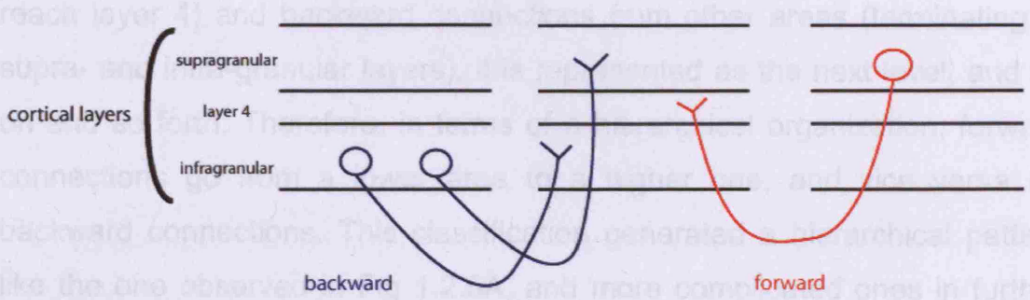


Fig 1.2.4. Forward and backward connections in the cortex. Forward connections originate in supragranular layers and terminate in layer 4. Backward connections originate in infragranular layers and terminate in infragranular and supragranular ones.

These histological differences between layers also subserve functional ones, i.e. distinct computations will be performed in each layer, which will result in a different pattern of connectivity with layers within the same or different areas. A connection can be defined as forward or backward depending on the cortical layers of origin and target. Forward connections originate in supragranular layers, terminate in layer 4 and have sparse axonal bifurcations (Fig 1.2.4). Backward connections originate in infragranular layers and terminate in supragranular and infragranular layers (Rockland and Pandya, 1979, 1981; Salin and Bullier, 1995).

In general, forward connections are *driving*, they would elicit a response in the neuron, reflected by the fact that they target granular layers, whereas backward connections are in general modulatory, meaning that they will change the responses to other inputs, as they terminate in supragranular and infragranular layers (for a review see Callaway 1998; 2004).

Based on this classification of connections, Maunsell and Van Essen (1983) defined a hierarchy of visual areas, where a particular cortical area is going to be at an early or late stage of this hierarchy depending on the kind of connections it has with other areas. The primary visual cortex represented the base of the hierarchy, receiving only backward connections and sending only forward ones (although later hierarchical models include also inputs to V1 from subcortical areas). Since V2 receives inputs from V1 (those that reach layer 4) and backward connections from other areas (terminating in supra- and infra-granular layers), it is represented as the next level, and so on and so forth. Therefore, in terms of a hierarchical organization, forward connections go from a lower area to a higher one, and vice versa for backward connections. This classification generated a hierarchical pattern like the one observed in Fig 1.2.5A, and more complicated ones in further refinements (Fig 1.2.5B) (Felleman and Van Essen, 1991), but neither with this classification nor with other ones (i.e. defining the hierarchy based on the amount of granular cells in layer; Barone et al, 2000), it is not possible to construct a unique hierarchy of visual areas - there are violations of the rules to incorporate all areas, and there are several distributions that are as

optimal. This can be due to lack of knowledge about the connections between areas or maybe because the cortex is not organized in such a strict hierarchy.

A

B

Fig 1.2.5. Hierarchical organization of the visual system. Each module in the diagrams represents a visual area. The lower part of each diagram represents the bottom of the hierarchy and the upper part represents the top. **A**. Modified from Maunsell and Van Essen, 1983. **B**. Modified from Felleman and Van Essen, 1991.

As Felleman and Van Essen (1991) pointed out: *'The possibility that the visual cortex might operate by a strictly serial processing scheme can be ruled out just from knowing the multiplicity of connections per area and the near ubiquity of reciprocal connections. On the other hand, it seems highly unlikely that the visual cortex is a network that altogether lacks any distinction between processing levels."*

Nevertheless, some interesting facts were made clear after this classification: 1) cortico-cortical connections are generally reciprocal, where information can be processed in loops between and within brain areas; 2) each area is going to integrate top-down and bottom-up information to perform its specific computation and send, again, information through both kinds of connections: backward and forward; and 3) backward connections cross several levels, while forward ones are more restricted [a clear example of this fact is V1, which receives multiple backward connections from regions to which it does not directly project to, like FEF, IT and the auditory cortex, between others (Salin and Bullier, 1995; Barone et al, 2000; Falchier et al, 2002).

Fig 1.2.6. Latencies of the neuronal responses in several brain areas. FEF: frontal eye fields; LIP: lateral intraparietal; IT: inferior temporal; OFC: orbitofrontal cortex. Figure modified from Bullier, 2004.

Another way that researchers have used to show that information does not flow through the hierarchy in a strict way is to measure the latencies of neuronal responses in each visual area (Schmolesky et al, 1998; Nowak and Bullier, 1995). Responses are faster in V1 than in the other areas, but there is a big discrepancy between the latencies of further areas and their position in the visual hierarchy. In general, latencies are shorter in posterior areas compared to anterior ones (V1 < V2 < IT < OFC), which follows a hierarchical transmission of information, but they are much shorter in area V3 than in V2, even when the latter is generally located before in the hierarchy (Fig 1.2.6). Dorsal areas like V5 and FEF have shorter latencies than their correspondents in the ventral brain. This is a reflection of the amount of myelinization in their axons and of their function as well: V5 should keep track of the changes in position of a determined visual object and FEF needs to send signals to reorient gaze towards salient, and therefore environmentally relevant, stimuli.

To end this section, I would like to explain the choice of the forward and backward terminology in this thesis. The terminology most widely used in the neuroscience community refers to feedforward and feedback connections. But, in doing so, a functional label is ascribed to the connections and, as Friston (2005) points out: *"There is something slightly counterintuitive about generative models in the brain. In this view, cortical hierarchies are trying to generate sensory data from high-level causes. This means the causal structure of the world is embodied in the backward connections. Forward connections simply provide feedback by conveying prediction error to higher levels. In short, forward connections are the* feedback *connections."* This highlights the point that a connection will be providing feedback depending on how we think the brain works, this is the reason why I prefer to use the terms forward and backward since it does not impose a function and just refers to which stage in the hierarchy is transmitting information towards.

FUNCTIONAL SPECIALIZATION IN THE BRAIN

The concept of functional specialization in the brain, key in modern neuroscience's view, dates to the 19th century, with Pierre Paul Broca's (1861) studies of an aphasic patient who could understand language but could not utter any word other than 'Tan' (and hence was called by this name in the hospital). Post-mortem observations of the injuries in the brain of the patient 'Tan' lead Broca to propose that the ability to articulate speech was located in the left frontal lobe, making the first causal link between a brain lesion and a specific behavioural deficit. This groundbreaking conclusion presented at the Societe d' Anthropologie in April 1861, set a solid foundation for the future studies of brain functions. This was the idea that the phrenologists, lead by Frank Gall, failed to demonstrate due to a misleading approach – the belief that personalitiy and behavioural traits characteristics of a certain individual will result in more developed brain regions that will reflect in the cranial structure as bumps. Even Jean Pierre Flourens (1824), who used a more informative approach, the study of the effects of lesions in the brains of birds and mammals, could not arrive at this conclusion. He showed that the cerebral hemispheres are responsible for higher cognitive functions, that the cerebellum regulates and integrates movements, and that the medulla controls vital functions (such as circulation, respiration and general bodily stability). He was convinced that the cortex was responsible for sensation, movement and thought, but could not find a particular localization of these functions. This led him to believe that they were represented in a diffuse form around the brain.

The evidence presented by Broca was soon supported and reinforced by the experiments of Gustav Fritsch and Eduard Hitzig (1870), who showed that stimulation of anterior parts of the cortex resulted in movements in the contralateral face and limbs, but this was not the case when the stimulation was delivered to posterior parts.

Broca's findings, and then Fritsch and Hitzig's confirmation, motivated scientists to the search of the anatomical loci of other functions. Of great relevance to the neuroscience world was the work of the German neurologist and anatomist Korbinian Brodmann (1905), whose histological studies of the brain led him to divide the cortex into fifty-two distinct

cytoarchitectonic regions, many of which are still in use and referred to as Brodmann areas (Fig 1.2.7).

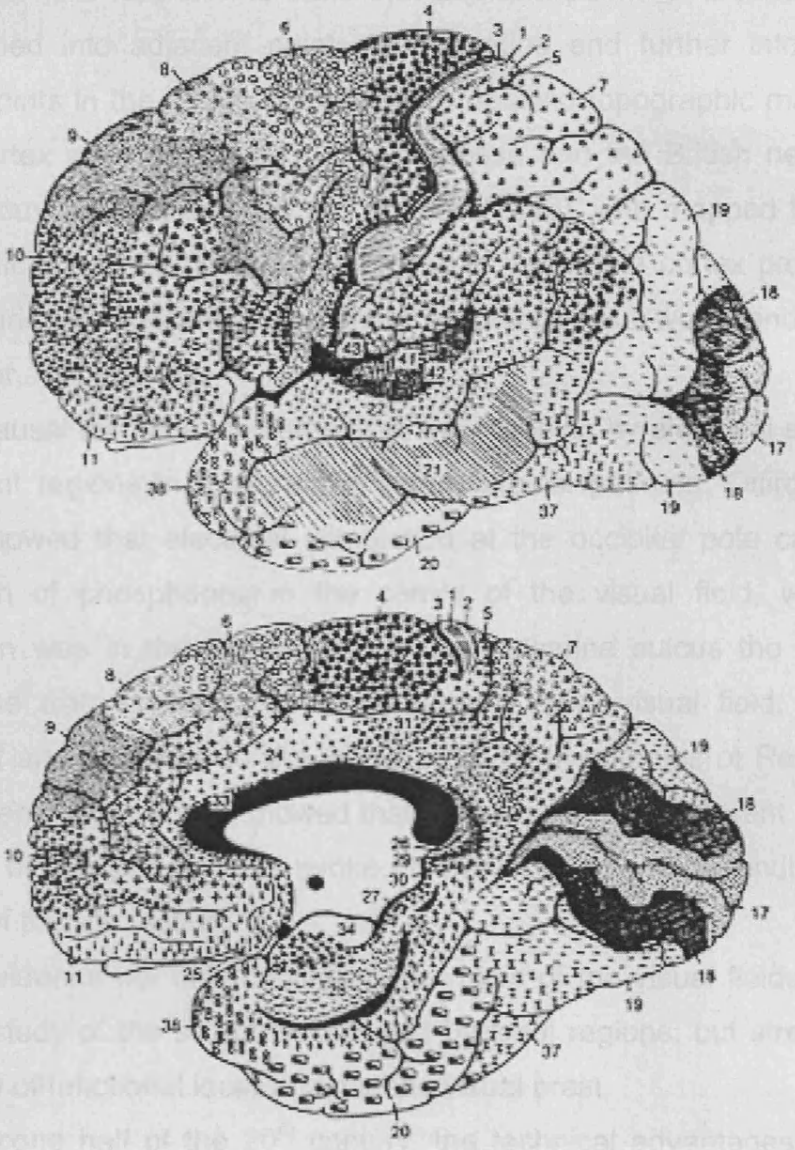


Fig 1.2.7. Brodmann areas.

In the study of vision, the work of Hermann Munk (1881) was indispensable for determining the role of the occipital lobe in this function. His experiments revealed that lesions in one occipital lobe made dogs and monkeys hemianopic, and if both hemispheres were lesioned the animals were completely blind, showing for the first time evidence for a topographic representation of the visual scene in the cortex. Further studies by Salomon Henschen (1893; 1900) on post-mortem brains of patients which had been

hemianopic due to strokes, revealed that the upper bank of the calcarine sulcus receives inputs from the upper retina and therefore from the lower visual field. This lead him to state that adjacent points of the visual scene are mapped into adjacent points in the retina and further into adjacent cortical points in the striate cortex. More detailed topographic maps of the striate cortex were provided by the Japanese and the British neurologists Tatsuji Inouye (1909) and Gordon Holmes (1945), who mapped the spatial visual deficits and the location of lesions in the striate cortex produced by bullet injuries in soldiers during the Russo-Japanese world and the First World War, respectively.

Further causal confirmation came from the mapping by electrical stimulation of different regions in the brain of neurosurgical patients. Otfird Foerster (1890) showed that electrical stimulation at the occipital pole caused the perception of phosphenes in the centre of the visual field, whereas if stimulation was in the upper bank of the calcarine sulcus the perceived phosphene was located in the contralateral lower visual field. This was confirmed and extended to the whole brain by the studies of Penfield and Rasmussen (1968), which showed that the stimulation of different regions of the brain of epileptic patients evoke different sensations depending on the location of the stimulation.

All this evidence not only confirmed the maps of the visual fields obtained with the study of the scotomas induced by focal regions, but strengthened the notion of functional localization in the visual brain.

In the second half of the 20th century, the technical advantages achieved during the war years allowed scientists to study the actual stimulus preferences of individual neurons by making *in vivo* extracellular recordings. It was the work of physiologists Torsten Wiesel and David Hubel at John Hopkins University that revealed the basic responses of cells in the visual cortex. Hubel and Wiesel's (1965, 1968) seminal papers describe how cells in the visual cortex are orientation selective – they respond to lines oriented in a particular angle falling into its receptive fields. They described three basic groups of cells: i) simple cells, with receptive fields (RF) that have spatially distinct 'on' and 'off' areas; ii) complex cells, that are always responsive to a defined orientation of a line stimulus if it is in their RF; and

iii) hypercomplex cells, subsequently called end-stopped cells, where lines extending beyond the RF cause a fall off in the response. They were the first who postulated a hierarchical organization of the visual brain, where there is an increase in the complexity of the computations performed from stage to stage, a knowledge supported by the increase in the number of complex and hypercomplex cells from V1 to V3. Hubel and Wiesel proposed that simple cells will receive inputs from centre-surround cells of the lateral geniculate nucleus, creating their particular RF characteristics, and then several simple cells will project to a complex cell, creating a hierarchical building-up process within and between areas (Fig 1.2.8).

Fig 1.2.8. Orientation selectivity of a simple cell. Several geniculate cells (blue), with 'ON' centres (+) arranged along a straight line in the retina, project upon a simple cell in the cortex (pink). As a result, the cortical cell has an elongated 'ON' receptive field (pink dashed-line), selective for a given orientation. Modified from Hubel and Wiesel, 1962.

In the 1970's, the work of Semir Zeki revealed the properties of areas beyond the striate and prestriate cortex and set the foundations of functional specialisation in the visual brain. Zeki showed that neurons in area V5, located in the posterior bank of the superior temporal sulcus, were selective

for motion stimuli (only responsive to this feature) (Zeki, 1974). On the other hand, neurons in area V4, located in the prelunate gyrus, contained neurons responding to colour stimuli, but not to motion (Zeki, 1973; 1978c). His work reinforced the knowledge that there are multiple areas outside the primary visual cortex, that the retina was independently represented in each and that the nature of retinal representations differs from area to area, which finally led him to postulate that the visual system is organised into multiple, parallel, functionally specialised subsystems.

With the introduction in the 1990s of Positron Emission Tomography for the study of brain functions, Zeki and collaborators showed, in the human brain, the presence of functionally specialised areas for the processing of both colour and motion, homologous to the areas he had discovered in the monkey (Zeki et al, 1991). Zeki (1993) had pointed out on many occasions that evidence for the functional specialisation of the human brain existed in the history of neurology. In particular, Louis Verrey's (1888) report of a patient with right hemiachromatopsia (the inability to see colour) after a stroke in the left occipital lobe and George Riddoch's (1917) studies of patients who had received gunshots during the First World War, who often reported seeing movement in their blind fields. These findings should have suggested the presence of colour and motion centres in the brain, but these theoretical implications were dismissed by the scientific community at the time, when the general view was that since there is a unitary perceptual experience of the visual world, the visual brain should also be processing the scene as a whole entity.

1.2.2 – PROCESSING OF VISUAL FORMS AND OBJECTS

VISUAL SCENES: THE BACKGROUND, THE FIGURES AND THEIR PARTS

When appreciating a visual scene, our experience is that of a unified whole, not that of millions of independent luminance points. We are capable of perceiving forms, of grouping them together into figures and of segregating

them from the background. The process by which the human brain organizes elements of the visual scene into wholes of different spatial scales is known as perceptual organization, and the one by which segmentation of a scene occurs is commonly referred as figure-ground segregation.

Here, I review some of the evidence from the psychological research of perceptual organization and figure-ground segregation. How these processes relate to neural functions will be reviewed in further sections.

The Gestalt group of psychologists was probably the first to have realized many of the automatic phenomena in the perception of a visual scene and to postulate some rules of perceptual organization that could lie behind them. The Gestalt school started in Germany around 1910, with Max Wertheimer's initial experiments on visual illusions like "apparent motion". This kind of phenomenon, at that time, was generally disregarded as a cognitive epiphenomenon. The novelty in the Gestaltist approach was to consider the overall visual scene (or a few "wholes"), instead of considering it the result of the combination of all the individual parts. This is clearly stated in their motto *"the whole is different from the sum of its parts"* and as Wertheimer wrote - *"The fundamental "formula" of Gestalt theory might be expressed in this way: There are wholes, the behaviour of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole. It is the hope of Gestalt theory to determine the nature of such wholes."*

Wertheimer started studying arrangements of dots and tried to understand the rules behind the grouping of one to another. In 1923 he formally announced some factors that determine perceptual grouping: proximity, similarity, common fate ("uniform destiny") and closure (See Fig 1.2.9 for a detailed description of each one). In his paper, Wertheimer also mentioned past experience as one of the factors that influence grouping. This factor relates to the history of the observer and not to the physical properties of the stimuli (a point I will explain in more detail below). In further studies during the first half of the 20th century, symmetry (Bahnsen, 1928) and parallelism (Morinaga, 1941) were also identified as perceptual grouping rules, and in the 90s, Stephen Palmer and collaborators proposed that elements are also grouped together if they are: a) located in the same closed region of space

(common region; Palmer and Rock, 1994); and b) connected by other elements (element connectedness; Palmer, 1992) (Fig 1.2.9).

Extending his studies, Palmer (1999) noted that some principles of grouping, such as proximity or similarity, resulted in *element aggregations*, where each component is still perceived as an individual element, despite the grouping of elements into a whole. On the other hand, rules like element connectedness and common fate (in some cases), result in *unit formation*, where elements are perceived as parts of a single unified object. Another automatic and indispensable process in the perception of wholes and parts is the segregation from the background. The Danish psychologist Edgar Rubin (1921) was the first to study rules behind the assignment of figure and ground in the scene. He realized that figures seem closer to the observer than the background and that figures are the ones that have the shape of the contour (i.e. shape is never assigned to the background). This has profound ecological justifications, because whereas the shape of a figure is intrinsic to it, any shape the background acquires is an accidental outcome of a specific arrangement of objects and the specific conditions in which they occur. It therefore makes sense that shape is assigned and analysed for figural regions only, and the background totally ignored (Fig 1.2.10A). In cases where a contour could represent two figures, observers will only perceive one side of the contour as being the figure at any given moment. This is obvious in the famous figure of Rubin's vase (named after the psychologist, although already known for a long time before him) (Fig 1.2.10B). Out of these observations and some further work, rules of figure-ground segregation emerged – a region of space will be more likely to be perceived as a figure if it is:

- totally surrounded by another region (Rubin, 1921).
- smaller than another region (Rubin, 1921).
- vertically or horizontally oriented (Rubin, 1921).
- highly contrasted (Rubin, 1921).
- symmetric (Kanizsa and Gerbino, 1976).
- convex (Metzger, 1953; Kanizsa and Gerbino, 1976).
- composed of parallel edges (Metzger, 1953).

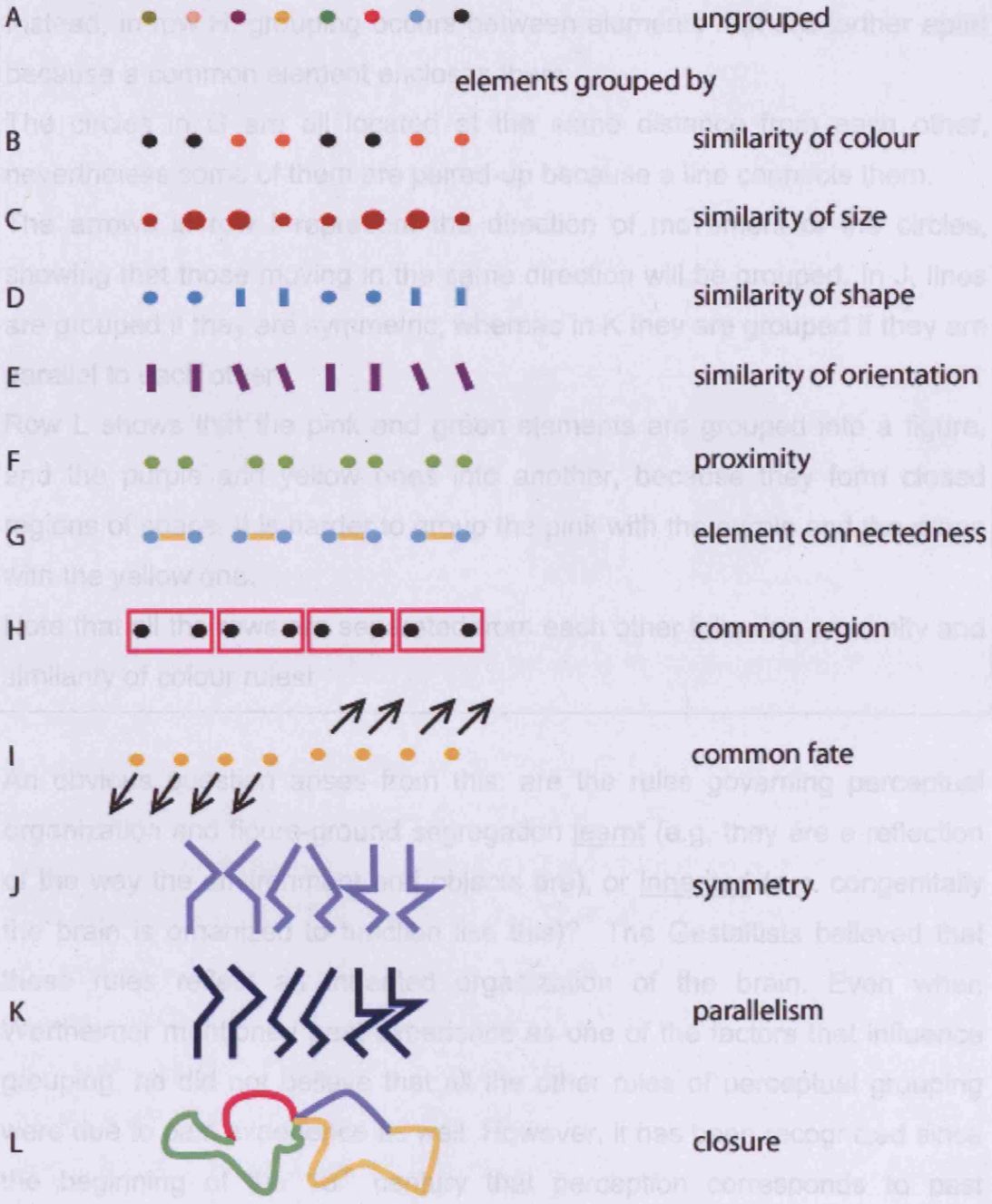


Fig 1.2.9. Perceptual grouping. Schematic examples of perceptual grouping rules. In row A, there is no automatic grouping of subsets of circles because they all have different colours. Instead, from B to E, the elements are grouped by similarity of features. Therefore, element 2 is grouped with element 1 in B, whereas the same element is

Fig 1.2.9. Perceptual grouping (*cont.*). grouped with element 3 in C. In row F, those circles that are closer to each other are grouped together. Instead, in row H, grouping occurs between elements that are farther apart because a common element encloses them.
The circles in G are all located at the same distance from each other, nevertheless some of them are paired-up because a line connects them.
The arrows in row I represent the direction of movement of the circles, showing that those moving in the same direction will be grouped. In J, lines are grouped if they are symmetric; whereas in K they are grouped if they are parallel to each other.
Row L shows that the pink and green elements are grouped into a figure, and the purple and yellow ones into another, because they form closed regions of space. It is harder to group the pink with the purple and the green with the yellow one.
Note that all the rows are separated from each other following proximity and similarity of colour rules!

An obvious question arises from this: are the rules governing perceptual organization and figure-ground segregation learnt (e.g. they are a reflection of the way the environment and objects are), or inherited (e.g. congenitally the brain is organized to function like this)? The Gestaltists believed that these rules reflect an inherited organization of the brain. Even when Wertheimer mentioned past experience as one of the factors that influence grouping, he did not believe that all the other rules of perceptual grouping were due to past experience as well. However, it has been recognized since the beginning of the 18th century that perception corresponds to past experience more than to the physical properties of a scene. In his work "An Essay Towards a New Theory of Vision", George Berkeley (1732) discusses that many of the judgments about distance, magnitude, and situation of objects are largely based on past experience. This is a very important observation, since the light that reaches the retina from any point in the visual scene is the result of a combination of properties such as reflectance, luminance and transmittance. Therefore, for any given spot of light there are many possible combinations of factors that could

Fig 1.2.10. Figure-ground segregation. A. i) An arrangement of figures in a white background – note that shape is only assigned to the figural elements, not to the background. ii) Some of the shapes that could have been assigned to the background (regions 1 and 2 in figure i). iii) The reorganization of the figures shows that shapes in panel ii no longer exist in this display, whereas the shape of the objects is preserved. This is why, ecologically, it makes senses not to assign shape to spaces in between figures, since it will be a waste of resources. Modified from Rubin, 2001. **B.** The classic Rubin vase-face. **C.** When two regions of space are separated by an edge, the one that has the shape of a known object more likely is going to be assigned as the figure, showing the influence that previous knowledge has on the segmentation of the scene. Modified from Petersen and Skow-Grant, 2003.

have caused it, which makes the information arriving at the retina profoundly ambiguous. A way of solving this problem is to use previous knowledge of the environment to constrain the variety of explanations. This

is what allows us to disambiguate the ambiguous information reaching the retina, but it is also the source of many striking visual illusions (Fig 1.2.11) (for contemporary reviews on this topic see Gregory, 1997; Purves et al, 2001; Howe and Purves, 2005; and see Bayesian models in the next section).

Furthermore, it has been shown that, in six-months old infants, the expected trajectory of an object corresponds to what they have learnt, and not necessarily to a straight line, which is what an adult would expect (Kochukhova and Gredeback, 2006). The rapid learning and long-term retention of new rules shown by six-month-olds suggest that the pre-existing assumption that objects move linearly could easily be derived from frequent exposure to such trajectories in the environment (Rakison, 2007).

The assignment of regions of space as figures or background is also influenced by previous knowledge. Mary Peterson and her colleagues (see Peterson and Skow-Grant, 2003 for a review) have shown that, when judging images composed of black and white regions sharing an edge, observers are more likely to report seeing a figure on the side of the border where a well-known object was sketched (Fig 1.2.10C). This effect disappears if the whole configuration is inverted or if the edge is scrambled, where similar structural characteristics are kept but in this case become devoid of any meaning. Evidence from computer science also supports this view, since successful computer models of figure-ground segregation have only been achieved when top-down knowledge is included to help the segmentation process (Ullman, 2006; Sharon et al, 2006).

If the rules of perceptual organization and figure-ground segregation are inherited or learnt is still open to debate, what seems to be obvious from all the evidence above is that there is a close relation between processes that determine the basic structure of the different components of a scene and the effects of beliefs and stored concepts. The variety of factors that determine grouping and figure-ground segregation suggest that these processes could occur at several stages of the visual brain, from very early ones, based on similarity of features, to late ones based on semantic knowledge. Considering the above evidence, it seems that segmentation

Figure 1.2.11. Visual illusions due to prior knowledge assumptions. A. A variation of the Cornsweet effect (Cornsweet, 1970), where the upper block of the central object is perceived as being darker than the lower one, but in fact they are physically identical. This happens because all the information in the stimulus is consistent with the adjoining territories being differently reflective (Image by Beau Lotto). **B.** The figure shows a Charles Chaplin mask in the upper left quadrant, with the nose and cheeks sticking out. If the mask is rotated (lower right quadrant), instead of perceiving a hollow mask, we still see a face with sticking out nose and cheeks. This is because we know that faces are volumetric figures, not hollow ones, even though the assumption is false in this case (Image by Richard Gregory).

and recognition do not occur independently but help each other through bottom-up and top-down loops as will be discussed in further sections.

THEORIES AND MODELS OF FORM AND OBJECT RECOGNITION

Object recognition is an automatic process for humans, but computationally it is a very difficult task. Recognition occurs at two levels: 1) identification, where an object is recognized as a unique exemplar and 2) categorization, where exemplars are recognized as being part of a general class. Therefore, recognition requires, in the case of identification, generalization

across different viewpoints, sizes, partially occluded parts and luminance conditions; and for categorization it also requires generalization across different exemplars of a class.

Two kinds of general approaches have been used in theories and models of visual processing, and therefore object recognition:

- Feed-forward models, where, in general, there is an orderly sequence of increasingly abstract information bearing an increasingly close correspondence to the three-dimensional geometry of the world (i.e. contours are grouped into surfaces and these into volumetric shapes). There are two main categories of this kind of model: 1) the classical compositional approach, where objects are represented as descriptions of spatial arrangements of parts in a three-dimensional coordinate system centred on the object itself; and 2) the image-based approach, where objects are represented as collections of view-specific features and recognition depends on previously seen object views.
- Top-down models, which in general include architectures that perform recognition by an analysis-by-synthesis approach: using a variety of image cues in a way that is simply dictated by what works, the system makes a guess about what object may be in the image and its position and scale. It synthesizes a neural representation of the object relying on stored memories and then measures the difference between this and the actual visual input. This difference is then used to correct the initial hypothesis.

This classification, while useful as a starting point, is very broad, and does not consider the overlap that exists between the strategies used in a particular model (i.e. many of the top-down models are based on a hierarchical system).

In the following pages I describe some of the most influential models and theories of form processing and object recognition and their theoretical implications. The experimental evidence that supports them will be discussed in further sections.

Hubel and Wiesel (1965) posited the first model of form processing, after their description of orientation selectivity in the visual cortex of the cat. This

model did not tackle object recognition as itself, but described how relatively simple information entering the retina could be converged into more complex representations in the primary visual cortex. They proposed that simple cells receive inputs from centre-surround cells of the LGN, creating their particular RF characteristics, and then several simple cells project to a complex cell. As a result, a hierarchical building-up process is created within and between areas, where each stage integrates simpler inputs into more complex outputs.

David Marr used Hubel and Wiesel's discoveries in his structural description theory (Marr and Nishihara, 1978; Marr and Hildreth, 1980; Marr, 1982), one of the most influential theoretical advances of the 20th century in terms of processing of forms and object recognition, where the goal was to reconstruct a faithful description of the objects and their spatial relations. Marr suggested that the visual system generates a sequence of increasingly symbolic representations of a scene, progressing from a 'primal sketch' of the retinal image, through a '2.5D sketch' to simplified 3D models of objects. He proposed that information from cells tuned to different spatial frequencies (or scales) is combined into 'tokens' that are likely to correspond to real-world entities such as an edge. The tokens comprising the primal sketch were then used as input for further processes such as object recognition. Therefore, Marr's proposal was also strictly hierarchical in nature, with successive forward steps of low, intermediate and high level stages of visual processing, which in the visual cortex could likewise be represented by areas that perform low, intermediate and higher level operations. He recognized that not all computations were forward, although he seemed to believe that low-level vision can operate independently of top-down influences.

In a further classical compositional approach model, Biederman proposed a "Recognition by Components" (RBC; Biederman, 1987) strategy. In RBC, recognition consists of extracting a view-invariant structural description of an object in terms of spatial relationships among volumetric primitives (*geons*). This description is then matched to stored descriptions. In this model, recognition will be view-invariant as long as the same structural description can be extracted from the different object views.

Another top-down approach is the one used in image-based models where, instead of creating a viewpoint-invariant structural description, objects and their parts are represented by a combination of stored two-dimensional views (Riesenhuber and Poggio, 1999; Edelman and Intrator, 2000; Ullman, 2006), in which multiple views have to be stored to represent all possible views of an object and exemplars of a category. In these models, recognition of new exemplars would be achieved by comparing them to the most likely stored combination. To explain image-based models I will use Ullman's feature-based model, which has both, bottom-up and top-down components. The reason for doing so is that this model contemplates an iterative and recurrent process for recognition and segmentation. This is relevant because, bottom-up segmentation applied to natural images is usually incomplete due to unavoidable ambiguities that cannot be resolved without prior knowledge of the object class. Therefore, this model represents a link between the bottom-up and top-down models. In the feature-based model, objects are represented by a hierarchy of fragments that are extracted during learning from observed examples. A repeated application of the feature-extraction process results in a hierarchical object representation of informative parts and sub-parts at multiple levels. A feature is considered informative if it reduces the uncertainty about the class. This will result in fragments that have class-specific features and are selected to deliver a high amount of information for categorization. The same hierarchy of fragments is then used for general categorization, individual object recognition and object-parts identification. Recognition is also combined with object segmentation, using stored representations triggered by the fragments to provide a top-down process that delineates object boundaries in complex cluttered scenes. Therefore, the hierarchical fragment representation will guide the segmentation process, intimately integrating segmentation and recognition. This has complementary advantages: the top-down process groups together image regions that belong to the same objects, despite region inhomogeneity and low-contrast boundaries, and the bottom-up process more accurately delineates the precise boundary locations.

Thus, in opposition to bottom-up models, top-down models consider object recognition as a statistical inference problem, where visual perception consists of finding a stored representation that most likely could have caused that retinal stimulation. This idea has been around since the times of the German physicist and psychologist Hermann von Helmholtz, whose seminal "Treatise of physiological optics" concludes: '*just as it is the characteristic function of the eye to have light/sensations, so that we can see the world only as a luminous phenomenon, so likewise it is the characteristic function of the intellect to form general conceptions, that is, to search for causes; and hence it can conceive of the world only as being causal connection.*"

Several researchers had formalized this idea using a Bayesian statistical framework, where decisions are based on the posterior probability distribution of a certain object, a concept I will explain below.

In Bayesian statistics, inferences are made taking into account not only the likelihood of an event (i.e. the probability of the event given the observed evidence; that particular observation data) but also the prior probability, which is the knowledge before the observations (i.e. the probability of that event itself; all the data from all previous observations). The combination of both these parameters gives the posterior probability of a certain event.

Therefore, if a certain image I is generated by, for instance, a vector of values S (like shape, viewpoint reflectance), it can be formalized in a generative model $S \rightarrow I$, where properties S generate the image I. With a Bayesian approach to perception, the model is inverted and the observer tries to estimate different variables $S_1 ... S_n$ (orientation, shape, colour, etc...) given an image I. Decisions are then based on the posterior distribution:

$$P(S|I) = P(I|S)\, P(S) / P(I)$$

Where P(S|I) is the probability of an object given an image, which is the result of the probability of that image given the object P(I|S) (i.e. the likelihood), the probability of the object p(S) (i.e. the prior probability), and the probability of the image p(I). The object with the highest posterior probability is then chosen. This implies that perception is a trade-off

between the reliability of the image features, represented by the likelihood, and the prior probability. If the image is very clear, decisions are going to be substantially based on the likelihood, but in ambiguous cases perception is going to be based on the priors (i.e. the beliefs) (Lee and Mumford, 2003; Kersten et al, 2004).

One of the theories that incorporates Bayesian inference in the brain is predictive coding, which is a good approach to understanding neuronal dynamics in relation to perception (Friston, 2002), since it does not need *a priori* knowledge of the underlying causes of the perception nor need to invert a model, which sometimes is not invertible.

In predictive coding the dynamics of units in a network are trying to predict the inputs. Instead of trying to find functions for the inputs that predict the causes they find functions of causal estimates that predict the inputs. It is a real time, dynamical scheme that embeds two concurrent processes: 1) the parameters of the generative or forward model change to emulate the real world mixing of causes, using the current estimates; and 2) these estimates change to best explain the observed inputs, using the current forward model. Both the parameters and the states change in an identical fashion to minimise prediction error. If the prediction matches the input, then the prediction error is less. Minimising the error is a way of maximising both: a) the likelihood of the input given that estimate; and b) the prior probability of the estimate being true. At any level of the system, the connection strengths of the model are changed so as to minimise the error between the predicted and observed inputs.

In the visual system, this will be accomplished by higher visual areas projecting their predictions about the stimulus to early areas via backward connections, where they will be subtracted from the incoming data (Rao and Ballard 1999; Friston 2003). Within this context, each stage of the form pathway may compute the discrepancy between the information about the actual visual scene coming from lower-level visual areas and predictions arriving from higher-level visual areas, which are based on previous visual experiences (Rao and Ballard, 1999). The reiteration of this process across successive levels of the cortex could therefore lead to more complex visual analysis.

Even when there is strong experimental and theoretical evidence that support a hierarchical organization in the visual cortex, not all the theories are based on a first forward transmission of information. For example, Moshe Bar (2003; 2004) proposed that a partially analysed version of the input image (i.e. a blurred image) is projected rapidly from early visual areas directly to the prefrontal cortex (PFC). This coarse image activates representations in the PFC of the most likely interpretations of the input image, which are then back-projected as an "initial guess" to the inferior temporal cortex to be integrated with the bottom-up analysis, which Bar assumes will be slower. The top-down process facilitates recognition by substantially limiting the number of object representations that need to be considered.

In the following sections I will discuss the experimental evidence of form and object perception mechanisms in the brain, obtained with electrophysiological, neuropsychological and imaging approaches, and how these data relate and support the theoretical models reviewed above.

THE FORM PATHWAY IN MONKEYS

The form pathway in the brain of the macaque extends from retinotopic areas in the occipital lobe to the inferior temporal cortex and frontal regions. Form is traditionally thought of as being processed in the ventral visual pathway (i.e. V1, V2, V4, IT), whereas the dorsal pathway would be involved in processing information about spatial location, movement and goal-directed actions (for examples see Milner and Goodale, 1993; Reddy and Kanwisher, 2007; Connor et al, 2007). Here I do not make that distinction because, as will be evident in this and further sections of the Introduction and Results, several dorsal areas also process form information.

Form processing is thought to begin in V1, where orientation selective cells extract information about local contours (Hubel and Wiesel, 1968). Orientation-selective cells have also been described in further areas, like V2, V3, V3A and V4 (Hubel and Wiesel, 1968; Zeki 1978b, c), making them suitable to process form information. In addition, these areas respond to more complex features. Recordings from cells in area V2 have been made in response to angles (Ito and Komatshu, 2004), illusory contours (von der

Heydt et al, 1984; Leventhal et al, 1998), border ownership (i.e. differential responses to a border depending on if it pertains to the right or the left part a figure) (Zhou et al, 2000) and combinations of these (Hegde and Van Essen, 2000). In V3, columnar organization is based on both orientation selectivity and stereoscopic depth (Adams and Zeki, 2001). In area V4, Zeki (1983) has shown wavelength- selective and orientation-selective cells, with many of the latter ones also having wavelength preferences and broader orientation tuning. Further studies in V4 have shown differential responses for border ownership and curvature (Pasupathy and Connor, 2001), where the response of the population will represent a specific shape (Pasupathy and Connor, 2002). Therefore, extrastriate areas respond to simple features (oriented lines), but also to some of intermediate complexity (angles and curvature).

In IT, where cells have larger RFs that sometimes include an entire hemifield (Gross et al, 1969), neurons respond to even more complex features (Logothetis and Sheinberg, 1996; Tanaka, 2000). Studying the responses of neurons in this area, Tanaka et al (1991) reduced the complexity of an effective visual stimulus in a systematic manner until the simplest pattern that would maximally drive the cell was determined. The degree of stimulus complexity required to drive a cell was found to increase, in general, from posterior regions in the IT (TEO) to anterior ones (TE). They also showed a modular arrangement, where cells responding to similar stimuli were closer to each other than those with different selectivity (Tanaka, 1993). Anterior regions of the IT also show stronger responses to 3D shapes, compared to more posterior regions, also supporting an increase in the complexity of the driving stimuli from posterior to anterior (Janssen et al, 2000a; b).

Most of the neurons are responsive to a specific view of an object, but a small population is view-invariant (i.e. they will respond to the same object independently of the view from which it is shown) (Logothetis et al, 1995; Booth and Rolls, 1998).

When comparing the responses of neurons in IT to forms and their component parts, Brincat and Connor (2004; 2006) showed that neurons respond to multiple parts (multiparts) with a linear combination of the

responses associated with each single part (the response to the object AB is the sum of the response to A and the response to B). Other neurons, by contrast, exhibit nonlinear selectivity for specific multipart configurations (AB more than A+B), the overall response being dominated by an interaction effect associated with the combined presence of two or more contour parts at specific positions. This nonlinear response has been argued to be consistent with a recurrent network process that effectively compares parts' signals across neurons to generate inferences about multipart shape configurations. Information about identity and category in the IT seems to be represented at the population level, as Hung et al (2005) demonstrated that accurate readouts can be obtained about these characteristics when analysing the responses of small populations (100 neurons).

The IT projects to the prefrontal cortex, entorhinal cortex, perirhinal cortex, the amygdala, the basal ganglia and several subcortical structures (see Logothetis 2000 for a review). In these areas, form information is going to be further processed into concepts and memories. In particular, Freedman et al (2003) showed that responses in prefrontal neurons reflect categorical judgments. Using a set of stimuli that parametrically changed between images of cats to those of dogs, prefrontal neurons show sharper between-category differences and lower within-category variance than those in the IT, showing that PFC responses are more categorical and those in the IT more related to the analysis of individual stimuli. It should also be mentioned that categorical judgments may not occur exclusively in PFC, since responses related to categorization of abstract stimuli have also been shown in cells of the lateral intraparietal area (LIP) (Freedman and Assad, 2006), and some fMRI evidence, discussed in the next section, also supports this knowledge. Independently of the contribution of further regions to the categorization of stimuli, there are further differences in the processing of form in these areas. Further studies have shown that responses to the same form stimuli remain unchanged in IT if the task's requirements change (Suzuki et al, 2006), but the opposite is the case for cells in the prefrontal cortex (Sakagami and Tsutsui, 1999; Sakagami et al, 2001). Therefore, even though the exact function of each of these areas still needs to be established, it seems that

the IT is processing more structural information about the image, while the frontal regions are processing more conceptual information.

Even when, at first glance, the visual pathways involved in the processing of form can appear as a straightforward hierarchical feature-extraction system, much evidence exists demonstrating that this is not so. Lamme has shown that in V1 responses to texture patterns are different if they are part of the object or if they correspond to the background (Lamme, 1995), showing that some higher-level form computations also take part in early areas. Since these differential responses are shown at latencies of 80 to 100ms, it is thought that they might reflect a top-down influence.

Responses of V4 neurons are modulated by selective attention (Reynolds and Chelazzi, 2004), at least in part mediated by signals from the frontal eye fields (Armstrong et al 2006; Armstrong and Moore, 2007), being able to selectively signal the behavioural category of the attended feature (Mirabella et al, 2007), showing that this area not only extracts structural information but is also sending a signal about an abstract categorical response. This response also occurs in the late part of the neural response, around 100-150ms.

In IT, responses of neurons are not only selective for structural information, but they also reflect learnt patterns. Learning of associations between different shapes (Sakai and Miyashita, 1991; Messinger et al, 2001) is reflected in IT cortical neural activity: neurons that show specific responses for one image also show an enhanced response if a stimulus learnt to be associated with that image is presented. It has also been shown that responses are enhanced for objects' parts which are highly diagnostic (i.e. an object having part "A" will have a big chance of being part of category X) (Sigala and Logothetis, 2004), being preferentially represented in the IT in posterior and anterior regions at single cell level, but only reflected in the local field potential at more anterior locations. Selectivity to learnt objects is not only enhanced for the whole shape, but also for the component parts. Here as well, the selectivity reflects a kind of conjunctive encoding whereby two parts together exert a greater influence on neuronal activity than predicted by the additive influence of each part considered individually (Baker et al, 2002).

The evidence mentioned above shows that form is processed in several areas of the visual brain, where there is an interplay of extraction of simple features and more abstract responses, like contextual, attentional and memory modulations, that influence the selectivity of the response and late components of it.

Single-cell recordings are in general obtained from neurons that are highly responsive to the tested stimulus. Even population responses are pooled out of all those responsive neurons, a procedure that ignores what happens to all those neurons that are silent (i.e. are they silent because of not being selective to the stimulus or due to a top-down influence?). This kind of process can be studied in humans using a variety of techniques like fMRI, EEG and MEG. In particular, it is interesting to see the fMRI results, since the BOLD response reflects in large part presynaptic activity (Logothetis et al, 2001), that in single cell recordings will not be taken into account if it does not elicit an output. In following sections I will discuss some of the evidence obtained with imaging techniques about mechanisms of form perception in humans, but first I will review the knowledge acquired through the study of patients with brain lesions that affect the perception of form. The neuropsychological approach has given deep insights into this problem for centuries, and has constituted a solid base for the study of these issues using imaging techniques.

THE NEUROPSYCHOLOGICAL APPROACH: AGNOSIAS

As described in the functional specialization section, the study of patients with brain lesions has been used for centuries to understand human brain function. In the case of form processing there are brain lesions that result in specific deficits in aspects of the normal perception of a form. These conditions are known as "agnosias".

The German neurologist Heinrich Lissauer (1988) was the first to publish a treatise on agnosia, distinguishing between two broad categories: "apperceptive agnosia" and " associative agnosia". The first refers to those conditions where deficits in the perception of objects are due to failures in the normal visual processing of forms and the second when objects are "seeing" normally, but they cannot be recognized.

Excellent reviews on different kinds of agnosias have been published (Humphreys and Riddoch, 1987; Farah, 2004), where the implications of different brain lesions on the processing of objects is extensively discussed. In this thesis, the aim is to show the kind of inferences that can be made about form perception using a neuropsychological approach. Therefore, from Lissauer's first classification it can be concluded that at least two processes take part in the perception of a form: the structural representation and then its association with semantic concepts. Lissauer's classification is still in use, mainly in neurology, although many types of agnosias that do not fit exactly in these groups have been reported, as will be evident in the following pages.

Patients with apperceptive or visual form agnosia have extrastriate lesions, often due to carbon monoxide poisoning, and show profound deficits in form perception – they cannot recognize, match, copy or discriminate simple visual stimuli (Adler, 1944; Efron, 1968; Benson and Greenberg, 1968; 1969; Milner and Goodale, 1993). Despite this, their visual acuity is within the normal range, with relatively unimpaired perception of colour, motion and stereopsis. Therefore, they have preserved perception of image features, but they do not succeed in grouping them. Remarkably, these cases also provide evidence in favour of the separation of visual functions in the brain.

In contrast, in associative agnosia the sensory system is well preserved, in the sense that patients can see and are able to report very accurately (generally by drawing) the visual scene. However, they cannot recognize objects either verbally or through other report modalities, like semantic grouping or mimicking of the function; showing dissociation between the structural description and the semantic meaning of a form. Object recognition through other modalities is unimpaired.

As a particular example I would like to mention Humphreys and Riddoch's patient HJA, who shows a special case of apperceptive agnosia, known as integrative agnosia, due to a ventral extrastriate lesion (Riddoch and Humphreys, 1987). This subject cannot recognize objects and has difficulties in everyday life, it takes him a very long time to make good drawings of the images he is presented, but he can nevertheless identify

simple shapes. This points towards a hierarchical system of object representations, where simple shapes, which are generally part of more complex objects, can be recognized, but not the more complex ones. It also suggests dissociation between the identification of local parts and a subsequent grouping into global configurations.

Agnosias can selectively affect the recognition of a particular class of objects. The most studied class-specific agnosia is prosopagnosia (Pallis, 1955; Farah et al, 1995; Kleinschmidt and Cohen, 2006; Berhrmann and Avidan, 2005; Duchaine and Nakayama, 2006), where subjects fail to recognize faces. In general, they have no difficulties in categorizing faces, but they cannot identify them and they often report seeing the parts individually, but not the whole. There are other class-specific agnosias where people cannot recognize buildings, city landmarks and streets, even when their spatial abilities are preserved, known as topographic agnosias (Landis et al, 1986; Mendez and Cherrier, 2003). Or cases where temporo-occipital lesions lead to pure alexia (ventral simultagnosia), where people can perceive objects without difficulty, but they are very impaired at reading and recognizing words (Dejerine, 1892; Warrington and Shallice, 1980; Patterson and Kay, 1982; Cohen et al, 2003; Kleinschmidt and Cohen, 2006).

These findings support the theory of some modules in larger areas involved in object recognition, either due to the congregation of neurons processing features which are highly present in these classes or due to specialization for a class, and they match the description of areas like FFA, PPA and the visual word area, where fMRI has revealed higher activations in response to faces, places and words.

There are also cases where subjects cannot perceive more than one object in the visual scene, known as simultagnosias (dorsal simultagnosias) (Luria, 1959; Farah 1990; Baylis et al, 1994; Gilchrist et al, 1996). They can identify single objects occupying the whole visual field, but they cannot identify multiple objects covering the same extent of the visual field. This seems to be an attention deficit problem, a hypothesis supported by the fact that subjects usually present parieto-occipital lesions and that the objects might seem to disappear even if neither the eyes nor the objects move. This is

probably due to habituation of central visual representations after prolonged gaze, typical of these patients after the allocation of attention onto a particular object. The study of simultagnosia shows that objects in a scene are processed separately, probably after edges have been assigned to each one, and suggests the possibility of competence between the representations of different object.

This summary shows that based only on the neuropsychological data, it is possible to understand many of the functions of the visual pathways involved in the perception of form. The combination of this approach with fMRI, EEG and MEG had been immensely valuable in understanding the human brain.

M/EEG AND fMRI EVIDENCE FOR OBJECT AND FORM PERCEPTION

The neuropsychological evidence reviewed in the previous section shows that several areas in the occipital and temporal cortices are involved in form processing and object recognition in humans. These findings have resulted in several hypotheses of how forms are processed in the brain. The use of techniques that measure neural activity in humans have allowed researchers to prove or falsify some of them. In the following sections I review some of this evidence and how it contributed to the current understanding of the brain mechanisms involved in the perception of forms and objects.

The first approach adopted in neuroimaging to study these issues was to compare activations elicited by objects and by non-object stimuli. The first brain imaging study of object perception was conducted by Haxby and collaborators (1991), with subjects performing a face-matching (object recognition task) and a dot-location matching paradigm (spatial location task). They showed that both visual-matching tasks activated the lateral occipital cortex, but face discrimination also activated a more anterior and inferior region of the occipito-temporal cortex and the spatial location task alone activated a region of the lateral superior parietal cortex, showing that also in humans parietal regions process spatial information and ventral occipito-temporal areas are related to object recognition.

In 1995, Malach et al described a region more responsive for object than textures in the lateral part of the occipital lobe which they called the lateral occipital complex (LOC). This term has been very influential, but refers to a very extensive portion of the cortex, which includes the lateral-posterior aspect of the occipital lobe, just abutting the posterior aspect of V5, extending towards the ventral occipito-temporal cortex anterior to V4, to posterior temporal regions and the anterior fusiform gyrus. It is therefore sometimes difficult to know which exact anatomical location researchers refer to when they use the acronym "LOC". In the studies I am mentioning, unless it is otherwise specified, LOC refers to all the voxels that are more active for objects compared to scrambled objects or non-objects in the ventral occipito-temporal and lateral occipital cortex. Notice that LOC is therefore mainly a functional classification.

The LOC has the necessary characteristics of an object representation system. Activations have been obtained in this area in response to a great variety of objects, including faces, houses, natural and man-made objects (Sergent et al, 1992; Malach et al, 1995; Grill-Spector et al, 1999; Haxby et al, 2001). Responses to faces, letter strings, and numbers have also been obtained with electrodes chronically implanted on the surface of striate and extrastriate cortex of epileptic patients prior to surgery. Authors found a negative potential, N200, associated with the stimuli and found that category specific responses were organized in "modules" or "patches" that vary in size and location (Allison et al 1994; Allison et al 1999; McCarthy et al, 1999). The LOC also activates with objects independently of the cue that defines them, including luminance, texture, motion, stereo or colour (Grill-Spector et al, 1998a; Gilaie-Dotan et al, 2002; Kastner et al, 2000; Self and Zeki, 2005). Furthermore, it is the region where the integration of information provided by different cues defining a single object occurs. Using displays of moving coloured-dots that sometimes defined a shape either by colour coherence, motion coherence or both, Self and Zeki (2005) showed that more anterior-ventral areas of the LOC (in the ventral occipito-temporal cortex) are activated when objects are defined by colour and motion, compared to when they are defined by a single cue. They also demonstrated, using an adaptation paradigm, which allows for identification

of regions invariant to a certain stimulus features (Grill-Spector and Malach, 2001), that this area adapts to a specific shape, independently of the cue that is used to define it. This suggests the presence of cue-invariant populations of neurons that respond to a particular object or shape independently of the defining cue. Also in support of this knowledge, Kourtzi and Kanwisher (2001) showed an adaptation of responses in the LOC when objects have the same shape but different contours, but not when contours are identical and the perceived shape is different. It should be noticed that this study did not evaluate responses in other brain areas, which leaves open the possibility of this same response occurring in other visual areas. Anyway, it seems that shape is a strong cue in the activation of the LOC, since this region is also active when shape information is extracted by other sensory modalities, like touch and audition (Amedi et al 2002; James et al 2002; Amedi et al 2007).

Adaptation paradigms have also shown, more significantly in anterior visual areas, invariance to external viewing conditions that affect the appearance but not the identity of objects, like object size, location, and viewpoint (Grill-Spector et al 1999; Vuilleumier et al 2002). Lerner et al (2001) showed that responses are reduced in anterior regions of the LOC by scrambling objects into four segments, whereas reductions in posterior regions were observed when objects were scrambled into sixteen fragments, supporting a posterior and anterior dissociation as well. Therefore, a subdivision of the LOC (Grill-Spector et al, 1999) can be made into a dorsal-caudal region named lateral occipital (LO) (Fig 1.2.3), with responses more sensitive to basic features, and an anterior ventral one (pFs/Loa or VOT) (Fig 1.2.3), which seems to respond in a more complex, holistic and invariant fashion, like it seems to happen in the monkey IT.

With all this evidence the question, which is still in debate, is what is the nature of the functional organization and the computations that take place in the LOC, in particular the ventral occipito-temporal part of it. The two main approaches are the following (although see also Malach et al 2002; and Gauthier, 2000):

- Some authors (Kanwisher et al, 1997; Epstein and Kanwisher, 1998; Kanwisher 2000; Downing et al, 2001; Spiridon and Kanwisher 2002)

propose that the ventral temporal cortex contains a limited number of areas specialised for representing specific categories of stimuli. A region in the lateral fusiform gyrus have been identified in humans as been specifically involved in the perception of faces (fusiform face area, FFA) since it responds more strongly to passive viewing of faces than objects or other human body parts, and also shows differential activity during consecutive matching tasks performed on faces versus hands (Sergent et al 1992; Kanwisher et al, 1997). Other regions more responsive to places (parahippocampal place area PPA) (Epstein and Kanwisher, 1998) and body parts (extrastriate body part area; EBA) (Downing et al, 2001), have also been identified. According to these authors, the remaining cortex, which shows little selectivity for a particular category, will be subserving a general purpose mechanism for processing of objects.

- Other authors propose an 'object form topography' model (Ishai et al, 1999; Haxby et al, 2001) which suggests that the ventral temporal cortex has a topographically organized representation of attributes of form that underlie object recognition, this region being "featurotopic" in the same way that primary visual cortex is retinotopic. The authors propose that bits of information most characteristic of a category of objects cluster together, resulting in a region that responds maximally to that category. This region does not represent a module since activation in response to each category elicits a stable and distributed pattern of activity in overlapping regions of the occipito-temporal cortex.

At present, there is no conclusive evidence in favour of one or the other account, and further evidence is necessary to disentangle between these possibilities, but there is a consensus that certain regions of the ventral occipito-temporal cortex respond more strongly to certain categories of objects, either due to a modular organization or to groups of cells responding to features of a certain category being clustered.

So far, I have made an account of the responses to forms in higher visual areas, but what kind of computation is taking place in earlier visual areas?

Grill-Spector et al (1998b) suggested a hierarchical sequence of neural stages in object processing after comparing brain activity elicited by objects relative to scrambled images, where areas V3A and V4 showed sensitivity to objects in a retinotopically defined fashion and the LOC had the highest responses independently of the location of the stimuli.

Responses to global contours of collinear elements relative to random oriented ones, have been shown not only in occipito-temporal regions, but also in early visual areas (Altmann et al, 2003; Kourtzi et al, 2003b), suggesting that shape perception could involve multiple visual areas that may integrate local elements to global shapes independently or sequentially at different spatial scales. In another study that tested responses to random lines and 2D- and 3D-shapes, higher responses were observed with random lines in V1 and with 3D-shapes in the LOC (Murray et al, 2002). Differential responses in other retinotopic visual areas were not observed. Ban et al (2006) found that activity in retinotopic areas elicited by an arc in a region of the visual field was greater when such an arc was perceived to be part of a circle. Dumoulin and Hess (2007) also showed stronger responses to Gabor patches forming circular patterns compared to random configurations in areas V3 and V4.

In studying the pattern of brain activity elicited by grouping elements either by similarity or proximity, Han et al (2005 a, b) found that proximity causes higher activity in the calcarine cortex, accompanied by a positivity at 100ms in the ERP, whereas similarity induces a negativity at a longer latency, which is located in occipito-temporal regions, proximity grouping being taking part in earlier stages than similarity grouping.

Altogether these results show that activity in earlier visual areas is also modulated by the global configuration of the stimuli, not only by the component basic features.

All the above evidence deals with the processing of forms independently of recognition and semantic mechanisms associated to them, but clearly it is not the same to be able to see a form and being able to relate to a concept, as it was already seen for patients with associative agnosia (Riddoch and Humphreys, 2003).

One of the most common approaches to study recognition and semantic processing is repetition priming, a form of memory by which repeated stimuli enjoy perceptual advantages over novel stimuli. (e.g., Evett & Humphreys, 1981; Forster & Davis, 1984).

Improved stimulus identification has been linked to decreases in the haemodynamic response for repeated as compared to newly presented stimuli (see Henson 2003; Henson and Rugg, 2003 for a review). This has been proposed to reflect a sharper stimulus representation that allows faster identification, which is mainly supported by the fact that priming-related decreases in brain activity are observed in category-specific regions of the ventral occipito-temporal cortex (Henson, 2003; Schacter et al, 2004).

Vuillemier et al (2002) used repetition priming to assess multiple stages in the processing of visual objects, including recognition variables (semantics, previous knowledge) and basic-features ones (size and view-dependent representations). They have found posterior-anterior differences in object processing, as well as left-right asymmetries. Bilateral regions in the lateral occipital and posterior inferior temporal cortex showed repetition-related decreases in activity independent of previous object knowledge (non-sense and real items). Anterior fusiform regions showed repetition effects selectively for meaningful stimuli, whereas repetition priming for objects of the same category was only observed in the left inferior frontal cortex. Responses to repetition priming in the right fusiform were view-dependent, not like in the left fusiform, where priming generalized across viewpoints, indicating more abstract representations in this hemisphere. These findings suggest that the visual system has a more anterior selectivity for known-objects and that it produces view dependent representations of new and known-objects. Equally importantly, it shows semantic and recognition responses are more anteriorly represented – responses to meaningful objects in anterior fusiform and to categories in frontal areas. In addition to this, single-cell recordings in hippocampal neurons of patients about to have surgery demonstrate invariant representations of learnt object associations (i.e. a neuron will respond to a picture of a person, its name and other people related to them) (Quiroga et al, 2005).

Studies have also been performed to study the involvement of these areas in the actual recognition of an object, not only passive viewing. Using bistable figures, where the physical characteristics of an image are the same but the perception changes, there is an increase in the activity of the FFA when people see the bistable figure as a face (Kleinschmidt et al, 1998; Hasson et al, 2001). The amount of activation elicited by objects in visual areas correlates with the level of recognition of an object: signals were higher for objects presented for longer presentation times and shorter masks or for those with the same timing before and after training (Fig 1.2.12) (Grill-Spector et al, 2000; Bar et al, 2001). In addition, differences between the levels of recognition are more pronounced in anterior visual

Fig. 1.2.12. Activity in visual areas correlates with recognition performance. **A.** Correlation between object recognition and activity in visual areas as a function of the presentation duration. DF: dorsal foci; LO: lateral occipital; pFs: posterior fusiform; CoS:collateral sulcus. Modified from Grill-Spector et al, 2000. **B.** The figure shows an increase in activity in the right fusiform face area (rFFA) when subjects successfully detected or identified faces. Modified from Grill-Spector et al, 2004.

areas than in more posterior ones. This shows that object recognition is not a discrete phenomenon constrained to an area in an all or none fashion, but

it rather takes place by a gradual increase in activity and recruitment of further anterior areas.

In an EEG experiment, Thorpe et al (1996) showed that the visual processing that is necessary to categorise images has to occur before 150ms, which is the time where they observed a frontal component, correlated to categorization. This fast processing is observed in cases in which images are easy to categorize, where recognition can occur in a feed-forward manner. But when the task is more difficult, due to short presentations, degradation of the stimuli or masking, there is a top-down influence from frontal and parietal areas that helps recognition (Bar et al, 2006; Eger et al, 2007).

Part 2 – METHODOLOGY AND RESULTS

The experiments presented in this part of the thesis investigate the neural mechanisms underlying form perception in humans. The evidence discussed in the Historical Survey sets a theoretical framework for the study of this subject. Visual perception in general, and form perception in particular, are the result of highly demanding and sophisticated computations. Sensory information arrives at the cortex, and here structural information about parts of the scene is extracted (i.e. the assignment of edges, the grouping of parts, segmentation) and associated to learnt concepts. We know that visual information about form is processed in a segregated manner in functionally specialised areas of the brain. However, the functions of each of these visual areas and how they interact with each other to make this sensory process successful are still not known. I approached this problem from three angles: I started with the study of the structural part of this process, the construction of a form, with an fMRI experiment designed to examine how the brain assembles all the component parts of a single form. Then I approached the semantic component of the perception of form, investigating regularities or known characteristics of images, and how different areas contribute to an explanation of current percepts through top-down and bottom-up mechanisms using fMRI and connectivity analysis. Finally, I investigated the evoked neural activity that subserve different levels of form recognition.

Each of these problems are studied in following chapters, where I introduce the subject, explain the methods used to investigate it and describe and discuss the experimental results. Some methods used in all the experiments are described in the first chapter (2.1) of this part. For information about general concepts behind fMRI, MEG, Retinotopic mapping and Dynamic Causal Modelling (DCM), the reader is referred to the appendices in Part 4.

Chapter 2.1 – GENERAL METHODOLOGY

Subjects

All the subjects that participated in the experiments of this thesis had normal or corrected-to-normal vision, were at least 18 years old and had no history of neurological or psychiatric disorders. They all gave informed written consent to participate in the studies, in accordance with the Declaration of Helsinki and with the approval of the Ethics Committee of the National Hospital for Neurology and Neurosurgery, London, UK.

Stimuli

All stimuli were designed and displayed using Matlab v6.1, v6.5 and v7.0 (Mathworks Inc.) and Cogent (www.vislab.ucl.ac.uk).

Imaging

BOLD contrast weighted echoplanar images (EPIs) were acquired with a 3T Siemens ALLEGRA scanner fitted with a head coil. Each volume consisted of a series of 38 axial slices of 2mm thickness and 1 mm gaps in between, with an in-plane resolution of 3x3x3mm, covering the whole brain with a repetition time (TR) of 2.47s. The first 5 volumes of each scanning run were discarded to allow for T1 equilibration effects.

T1-weighted anatomical images acquired in 176 slices of 1mm thickness, covering the whole brain plus cerebellum, were obtained for each subject.

In the display of all images, the right side of the brain is diplayed on the right side.

Chapter 2.2 - FORM CONSTRUCTION IN THE VISUAL BRAIN

INTRODUCTION

In a visual scene, any distinguishable individual pattern will have a form. To perceive a form as a whole, it is necessary to assemble together the simple spatially contiguous regions that constitute it, a process that we will refer to as form construction. This is a critical step in visual perception, fundamental for further object recognition and understanding of the environment.

The focus of this chapter is on the neural correlates of form construction. How these mechanisms then relate to the perceptual experience of individuals and lead to effects like global precedence (Navon, 1977) and configural superiority (Pomerantz et al 1977) should be the focus of further studies and it will be partially discussed in Chapter 2.3.

Electrophysiological studies of form perception in the visual brain of macaque monkeys have identified a pathway that includes retinotopically-defined areas in the occipital lobe (V1, V2, V3, V3A and V4) and areas in the inferior temporal cortex (IT) (Hubel and Wiesel, 1968; Zeki, 1978b; Logothetis et al, 1995; Tanaka, 2000). As one moves from V1 to IT, neurons have larger receptive fields (RF) and are responsive to more complex features (Hubel and Wiesel, 1965; Zeki, 1978c). In the light of this, form construction has been traditionally thought of as a forward hierarchical process, where each area, which is serially connected anatomically to the antecedent ones, will process more complex forms than the previous ones. In this scenario, cells with smaller RFs will respond to simpler components (oriented lines in V1)(Hubel and Wiesel, 1962; 1965; 1968; Zeki 1978c), cells in further areas such as V3, with larger RF, will process some intermediate components (angles, corners and curvature)(Dobbins et al, 1987; Versavel et al 1990; Pasupathy and Connor 1999; Hegde and Van Essen 2000; Ito and Komatsu 2004) and cells in yet further areas finally integrate the whole (shapes in IT) (Desimone et al 1984; Logothetis et al 1995; Tanaka 2000). From a theoretic point of view, classical compositional

models also support a modular hierarchical approach where information is transferred from one stage to the other in a bottom-up fashion (Marr and Nishihara, 1978; Biederman, 1987). However, a purely forward model cannot account for the sharp tuning in V1, which is accomplished by also integrating local lateral inhibition and recurrent cortical excitation (Somers et al, 1995), or for the responses of monkey's IT neurons to individual contour components and to combinations of them (multipart elements). Brincat and Connor (2006) showed that some neurons in IT respond to multiple parts (multiparts) with a linear combination of the responses associated with each single part (the response to the object AB is the sum of the response to A and the response to B). Other neurons, by contrast, exhibit nonlinear selectivity for specific multipart configurations (AB > A+B), the overall response being dominated by an interaction effect associated with the combined presence of two or more contour parts at specific positions. This nonlinear response has been argued to be consistent with a recurrent network process that effectively compares parts signals across neurons to generate inferences about multipart shape configurations.

From studies in human subjects using fMRI, a hierarchical sequence of neural stages in object processing has been suggested after comparing brain activity elicited by objects relative to scrambled images, where areas V3A and V4 showed sensitivity to objects in a retinotopically defined fashion and the LOC had the highest responses independently of the location of the stimuli (Grill-Spector et al, 1998b). However, responses to global contours of collinear elements relative to random oriented ones, have been shown not only in occipito-temporal regions, but also in early visual areas (Altmann et al, 2003; Kourtzi et al, 2003b; Ban et al, 2006) suggesting that shape perception could involve multiple visual areas that may integrate local elements to global shapes independently or sequentially at different spatial scales. In another study that tested responses to random lines and 2D- and 3D-shapes, higher responses were observed with random lines in V1 and with 3D-shapes in the LOC (Murray et al, 2002), arguing in favour of a predictive coding theory of neural processing (Rao and Ballard, 1999; Friston, 2003).

Based on this evidence, we picture three possible scenarios of form construction in the visual brain (Fig 2.2.1):

- A purely **hierarchical forward** mechanism, where each area will process simpler inputs into more complex outputs. This being the case, in the first stage simpler elements will be processed and this information sent to further areas, which will process more complex elements. A further stage will only get activated if the forms have the attributes to which their neurons are selective for. If this model holds, increments in form complexity will recruit further brain areas (Fig 2.2.1).
- The second scenario will be a **recurrent inhibitory** one, where each visual area preferentially responds to a particular level of form complexity. This could be achieved via backward and lateral connections that, by targeting inhibitory neurons, send inhibitory signals to previous stages or neighbour neurons. It can also be achieved through a mechanism that distributes signals to a particular brain area depending on the attributes of the image (in this case some extraction of the properties of the image will have to be made at the first stage) (Fig 2.2.1).
- A third model will be a **recurrent excitatory** one, where the activity in visual areas increases with the complexity of the form. Here all the activated stages contribute to finally construct the form through a recurrent and iterative process, where information is compared from one stage to the other until refined (Fig 2.2.1).

To study the mechanisms of form construction in the human visual brain, we carried out an event related fMRI experiment where subjects viewed stimuli with forms of different complexity levels.

The rationale behind our design was that the simplest elements of our stimulus set resemble those to which V1 cells are optimally responsive (oriented lines) and that the units of one group will be the constitutive elements of the next one (following the phenomenal experience of objects being composed of parts and subparts).

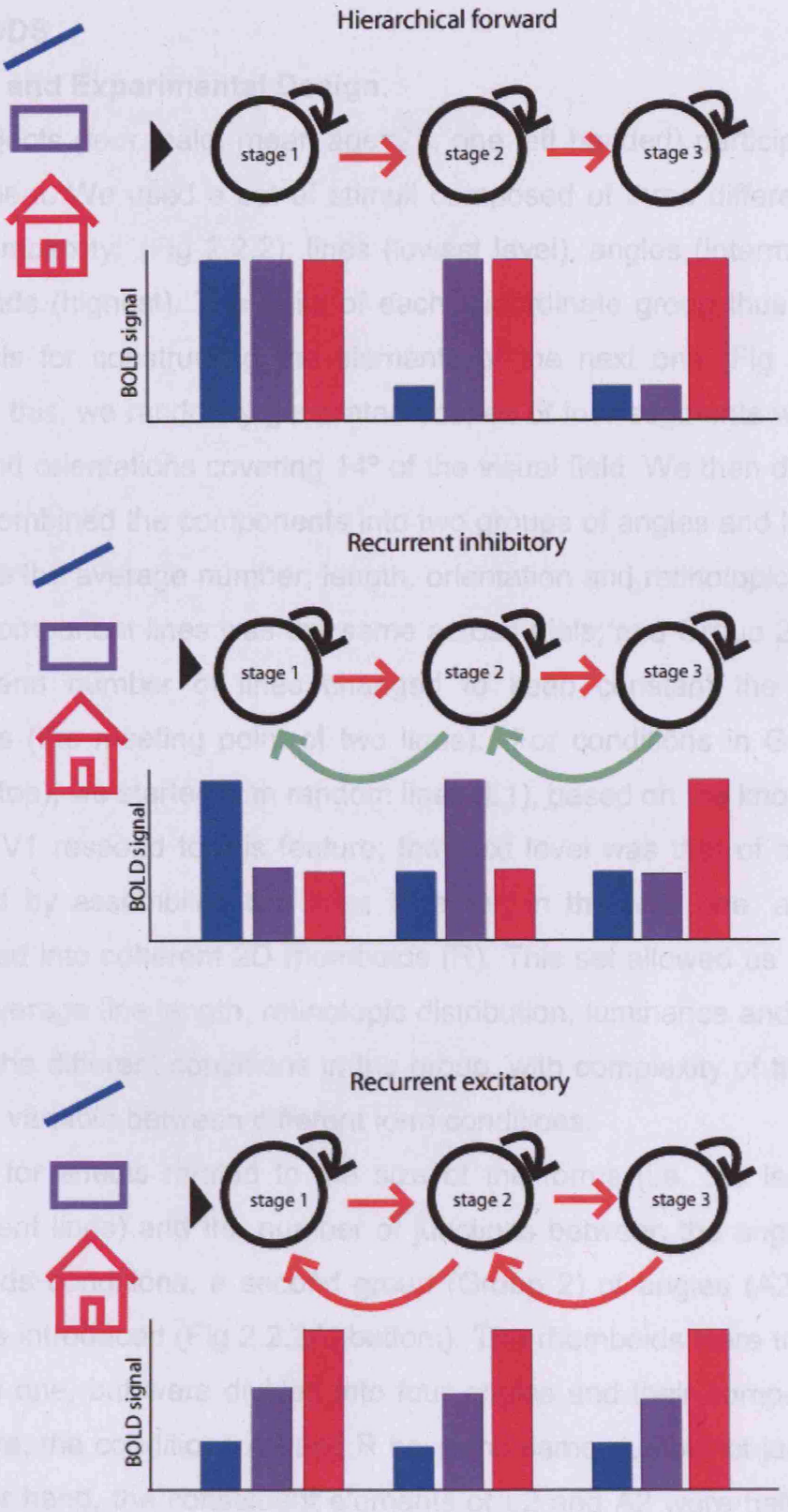


Fig 2.2.1. Schematic representation of some possible models of form construction. Each panel shows a hypothetical three-stage model and the BOLD response evoked at each stage by each form stimulus. Red arrows indicate excitatory signals and green arrows indicate inhibitory ones. See the text for details.

METHODS

Stimuli and Experimental Design.

Six subjects (four male, mean age=25, one left handed) participated in this experiment. We used a set of stimuli composed of three different levels of form complexity (Fig 2.2.2): lines (lowest level), angles (intermediate) and rhomboids (highest). The units of each subordinate group thus constituted the basis for constructing the elements of the next one (Fig 2.2.2A). To achieve this, we randomly generated shapes of four segments with different sizes and orientations covering 14° of the visual field. We then decomposed and recombined the components into two groups of angles and lines: Group 1, where the average number, length, orientation and retinotopic distribution of the component lines was the same across trials; and Group 2, where the length and number of lines changed to keep constant the number of junctions (the meeting point of two lines). For conditions in Group 1 (Fig 2.2.2A, top), we started with random lines (L1), based on the knowledge that cells in V1 respond to this feature; the next level was that of angles (A1), obtained by assembling two lines together; in the final one, angles were structured into coherent 2D rhomboids (R). This set allowed us to keep the same average line length, retinotopic distribution, luminance and orientation across the different conditions in the group, with complexity of the forms as the only variable between different form conditions.

To test for effects related to the size of the forms (i.e. the length of the constituent lines) and the number of junctions between the angles and the rhomboids conditions, a second group (Group 2) of angles (A2) and lines (L2) was introduced (Fig 2.2.2A, bottom). The rhomboids were the same as in group one, but were divided into four angles and their component lines. Therefore, the conditions A2 and R have the same number of junctions. On the other hand, the constituent elements of L2 and A2 were half the length (0.3-2°) of those in L1 and A1 (0.6-4°), making the comparison between groups of forms suitable for identifying size-related effects with forms of the same complexity. We used a wide range of size of elements to avoid constraining responses to a visual area with cells of a particular RF size.

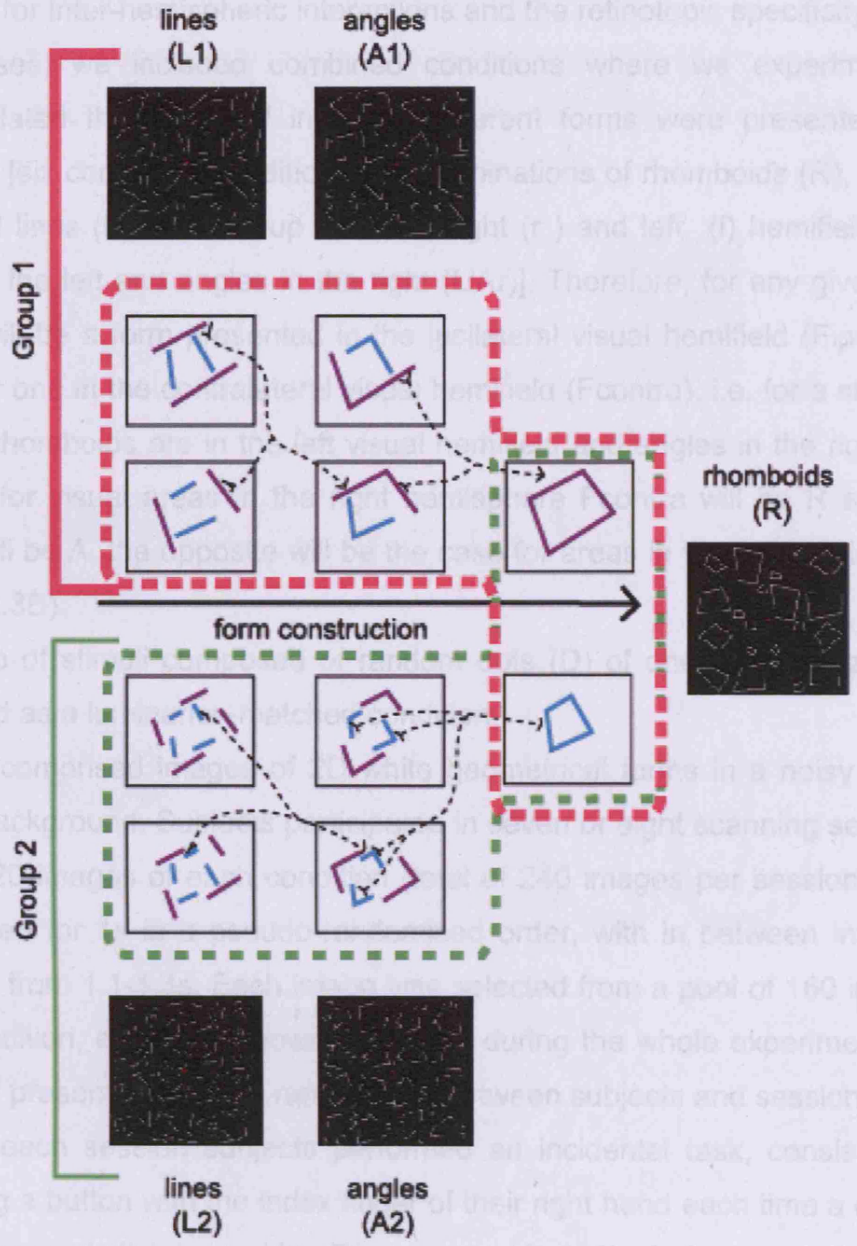


Fig 2.2.2. Stimuli. Examples of images from Group 1 and Group 2 and a schematic representation of the construction of the stimuli. Each group had three conditions: lines (L), angles (A), and rhomboids (R). Elements from one condition (i.e. L1) were combined to make the next one (i.e. A1). Conditions in Group 1 have, on average, components with the same line-length, orientation and retinotopic distribution. In Group 2, the lines (L2) and angles (A2) conditions have elements with line-length of half the size of R, but A2 and R have the same number of junctions. In both groups the R condition was the same.

To test for inter-hemispheric interactions and the retinotopic specificity of the responses, we included combined conditions where we experimentally manipulated the hemifield in which different forms were presented (Fig 2.2.3A) [six combined conditions, all combinations of rhomboids (R), angles (A) and lines (L) from Group 1, in the right (r) and left (l) hemifields; i.e. lines in the left and angles in the right (LlAr)]. Therefore, for any given ROI there will be a form presented in the ipsilateral visual hemifield (Fipsi) and another one in the contralateral visual hemifield (Fcontra), i.e. for a stimulus where rhomboids are in the left visual hemifield and angles in the right one (RlAr), for visual areas in the right hemisphere Fcontra will be R and the Fipsi will be A, the opposite will be the case for areas in the left hemisphere (Fig 2.2.3B).

A group of stimuli composed of random dots (D) of one pixel in size was included as a luminance-matched condition.

Stimuli comprised images of 2D white geometrical forms in a noisy (30%) black background. Subjects participated in seven or eight scanning sessions where 20 images of each condition (total of 240 images per session) were presented for 1s in a pseudo-randomised order, with in between intervals ranging from 1.1-1.3s. Each image was selected from a pool of 160 images per condition, each one shown only once during the whole experiment, the order of presentation being randomised between subjects and sessions.

During each session subjects performed an incidental task, consisting of pressing a button with the index finger of their right hand each time a central cross changed slightly in width. This occurred in 10% of all presentations.

Stimuli were viewed through an angled mirror on a screen located at a distance of 60cm onto which the stimuli were projected using an LCD projector.

Imaging

BOLD EPIs and T1-weighted anatomical images were acquired as described in the General Methodology section. Images were preprocessed and analysed using SPM2 software (**http://www.fil.ion.ucl.ac.uk/spm**). They were realigned to the first volume of the first experimental session, resliced to a final voxel resolution of 3x3x3mm and coregistered to each

subject's structural scan. They were then realigned in time as if every slice was acquired at the same time as the middle one (19th) of each volume and spatially smoothed with a Gaussian kernel of 6mm full width at half maximum.

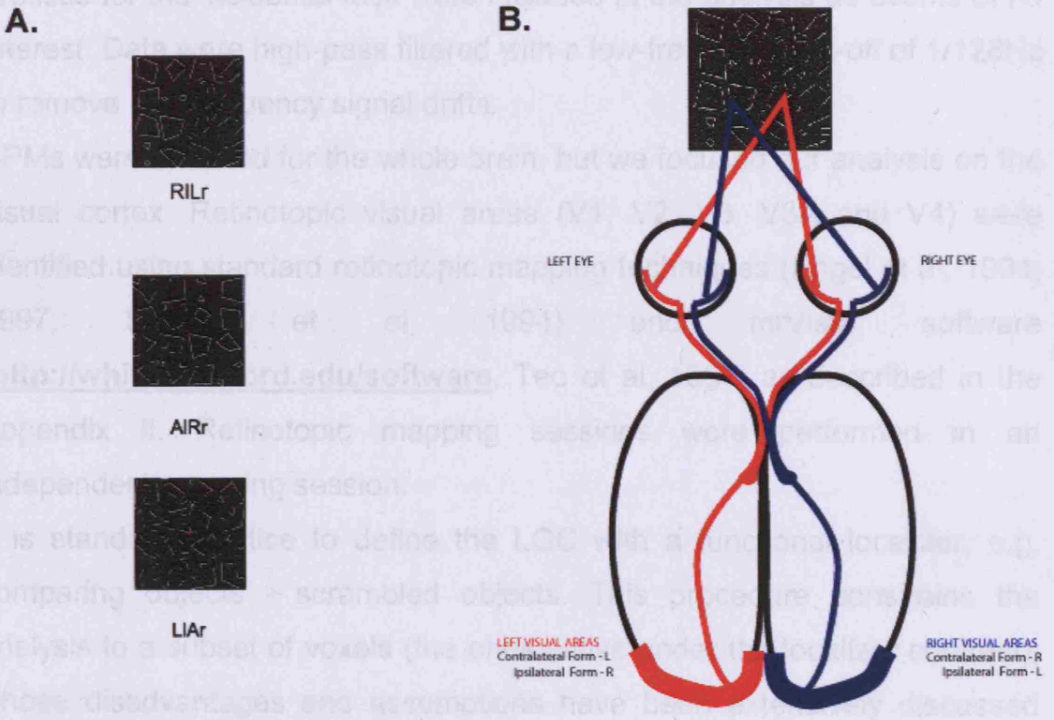


Fig. 2.2.3. Combined Conditions. A. Examples of images from the combined conditions. There were in total six combined conditions with all the combinations of lines (L), angles (A) and rhomboids (R) in the left (l) and right (r) hemifield (i.e. AIRr means angles in the left hemifield and rhomboids in the right one). A and L are the same as in Group 1. **B.** Schematic representation of the visual pathways activated by ipsilateral and contralateral form stimuli.

Main Analysis

The experiment was designed in an event-related manner with a total of 12 different conditions [dots (D), rhomboids (R), lines and angles from Group 1 (L1 and A1) and Group 2 (L2 and A2), and six combined conditions (RIAr, AIRr, AILr, LIAr, RILr, LIRr] each one modelled separately. Each stimulus presentation was modelled as a stick-function (a boxcar of duration 1/16th of

the TR), then convolved with SPM2's canonical haemodynamic response function (HRF) and entered into a multiple regression analysis to generate parameter estimates for each regressor at every voxel. The head movement parameters obtained during the realignment step and the subjects' button presses for the incidental task were included in the analysis as events of no interest. Data were high-pass filtered with a low-frequency cut-off of 1/128Hz to remove low-frequency signal drifts.

SPMs were obtained for the whole brain, but we focused our analysis on the visual cortex. Retinotopic visual areas (V1, V2, V3, V3A and V4) were identified using standard retinotopic mapping techniques (Engel et al, 1994; 1997; Sereno et al, 1994) and mrVista software (**http://white.stanford.edu/software**, Teo et al, 1997) as described in the Appendix II. Retinotopic mapping sessions were performed in an independent scanning session.

It is standard practice to define the LOC with a functional localizer, e.g. comparing objects - scrambled objects. This procedure constrains the analysis to a subset of voxels (the ones active under the localizer contrast), whose disadvantages and assumptions have been extensively discussed (Friston and Henson, 2006; Friston et al, 2006). In short, to reduce functional anatomy to a functional ROI (fROIs) assumes we know *a priori* the parcellation and segregation of function within the cortex. This is clearly not the case for the LOC, therefore we considered this approach as biased and based our analysis on purely anatomical definitions of regions. All voxels in the lateral surface of the occipital lobe, at the posterior part of the inferior-temporal sulcus, were labelled as LO; and those in the ventral occipito-temporal cortex anterior to V4, from the collateral sulcus medially to the occipito-temporal sulcus laterally, were labelled VOT. In this way, we are able to identify any voxel significantly active under the tested contrast, without restriction to the ones that are active for the localizing one.

Following the standard practice in the visual neuroimaging community, the percent signal change for each event was extracted from all voxels of each retinotopically or anatomically defined ROI with positive values for the contrast [R - D] using the MARSBAR toolbox for SPM2 (Brett et al, 2002; **http://marsbar.sourceforce.net**). Events were averaged for each condition

and each subject and expressed as percent signal change from condition D. Results from individual subjects were then used for group-level analysis using two-tailed t-tests and repeated measures ANOVAS where appropriate using SPSS software.

ROI analysis assumes that the voxels comprising the fROI (selected with a contrast) do not change with neuronal context or the level of task analysis. This is not necessarily so since a region that is activated in one context but deactivated in another will exhibit no main effect and will be missed using a constrained search procedure. Since the ROI analysis approach constrains the analysis to voxels more active for the R condition compared to the D as explained above, we also analysed the data on a single subject basis (Friston et al, 1999).

For the single subject analysis, SPMs were overlapped into 3D reconstructions of the cortex with delineated retinotopic areas for localization purposes. Mask volumes were created to identify significantly active voxels within a defined retinotopic area. For display purposes, the SPMs are threshold at $p<0.001$ uncorrected, but we only discuss results that were significant at $p<0.0001$ (i.e. $Z=3.7190$). The use of a $p<0.0001$ criteria provides a Bonferroni correction for 500 resolution elements (Worsley, 1994; Worsley et al, 1996). Given that we smoothed our data with a 6mm kernel, 500 resolution elements corresponds to 1413 mm^2, which is roughly the area of each visual ROI we examined.

To test for a linear increase in activity with an increase in form complexity, we conducted an independent parametric analysis (Buchel et al, 1998) where R, A1 and L1 were modelled as a single covariate with different weights for each condition. We followed a polynomial expansion strategy, modelling zero-, first- and second-order basis functions. The D, A2, L2 and the combined conditions, were modelled as parameters of no interest. Results were qualitatively the same when conditions of Group 2 were included in the parametric regressor. Parameter estimates for each parametric regressor and each subject were then used for group-level analysis using SPSS software.

Eye Movements

Eye movements were measured in a separate experiment outside the scanner. Six different subjects took part in three experimental sessions, with the exact same trial configuration and task as the one carried out in the scanner. The percentage of eye movements bigger than 0.5 degrees from fixation in shown in Fig 2.2.4. A repeated measures ANOVA showed no significant differences in eye movements between conditions [Y-axis: $F_{(11,55)}=1.69$, P=0.1; X-axis: $F_{(11,55)}=1.72$, P=0.09] nor significant deviations from fixation in the grand mean eye position for each condition ($F_{(11,55)}<1$, P=0.65), showing that fixation was well maintained during this experimental paradigm.

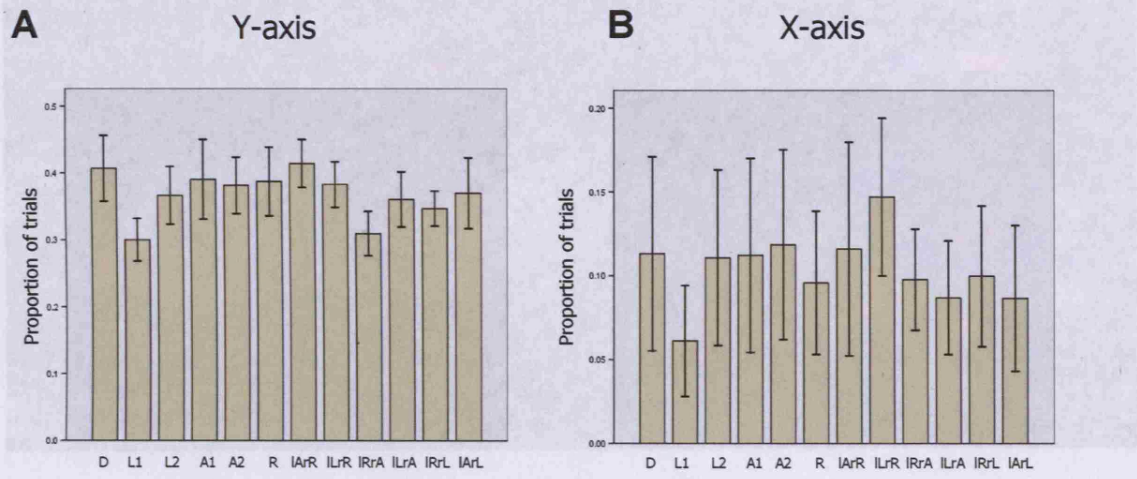


Fig. 2.2.4. Eye movements. A separate experiment was conducted outside the scanner to measure eye movements. The graphs show the proportion of trials in which eye movements were larger than 0.5 degrees from fixation in the Y (A) and X (B) axes. Bars represent mean ± s.e.m.

RESULTS

We began with a whole brain analysis of the contrast [R - D] and [D - R], to identify regions activated and deactivated by forms, respectively. Fig 2.2.5A and C show that significant activations are mainly located in the visual cortex, in agreement with previous results (Altmann et al, 2003; Kourtzi et al,

2003b); therefore we focused our analysis on retinotopic visual areas, LO and VOT.

Table 2.2.1 shows Z scores for all visual areas and all subjects for the maxima of the contrast [R - D] and [D - R]. With the single subjects analysis it is possible to appreciate that, even within a determined visual area, there

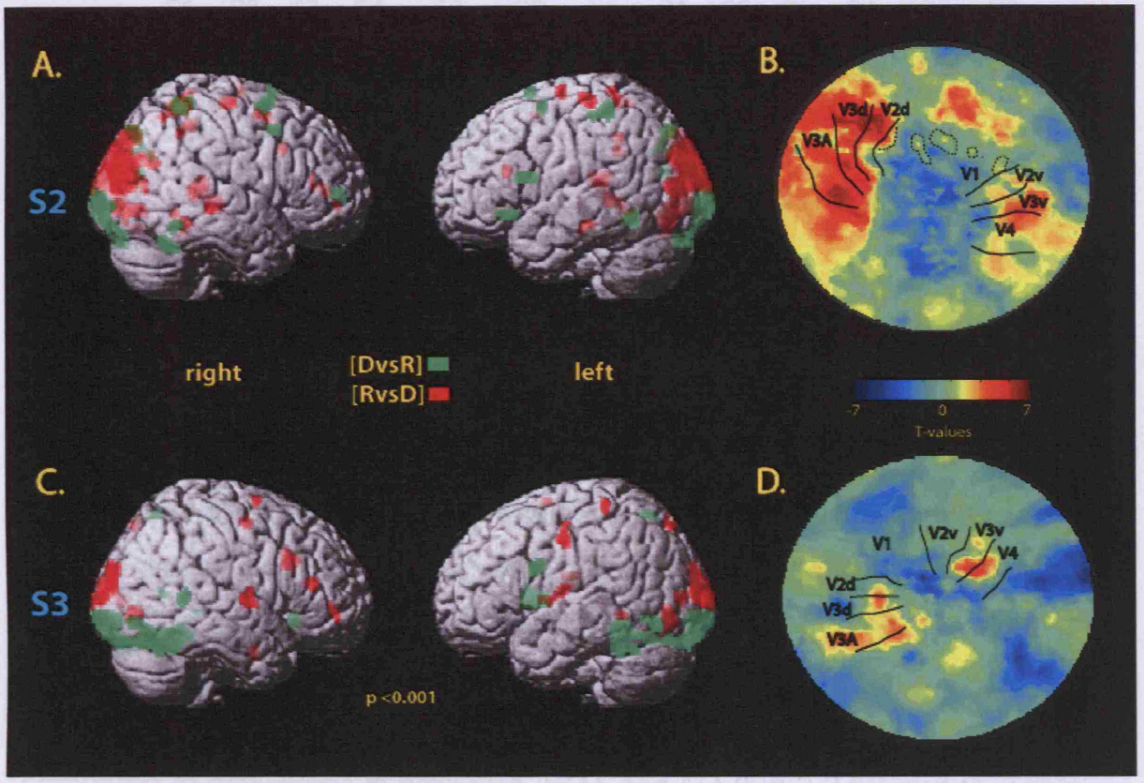


Fig 2.2.5. Form-related activations. A and C. The figure shows SPMs of a whole brain analysis of S2 and S3 for the contrasts [D vs R] (green) and [R vs D] (red) rendered into a normalised 3D brain. All voxels displayed survived a significance threshold of $p < 0.001$ (uncorrected) and an extent threshold of 15. **B and D**. SPMs for the contrast image [R vs D] overlaid onto flatmaps of the left occipital lobes of the same subjects. Colours represent T values (negative values represent deactivations). The dashed lines in B surround regions of activations (positive values) within V1.

can be subregions more responsive to different conditions. An analysis of the contrast [R - D] in each visual area reveals that activations with forms are more significant and consistent in areas V2, V3 and V3A

Table 2.2.1. Z-scores for the contrasts [R - D] and [D - R] at the single-subject level.

	[R - D]						[D - R]					
	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**
V1	**3.78**	**5.21**	2.11	2.57	3.65	3.67	1.91	**4.30**	**5.62**	3.26	**4.29**	**3.98**
V2	**4.21**	**6.74**	3.30	2.57	**5.60**	**3.97**	x	3.31	**4.03**	3.04	**4.35**	2.66
V3	**4.73**	**7.25**	**4.54**	**4.47**	**6.71**	**3.98**	x	3.00	3.26	2.28	2.49	2.29
V3A	**5.01**	**8.67**	**4.15**	**3.85**	**7.62**	**4.61**	x	x	x	2.66	1.97	x
V4	3.22	**3.94**	**4.46**	x	**6.25**	3.37	x	3.12	**4.84**	3.47	x	2.49
VOT	3.71	**3.72**	3.39	x	**3.90**	3.66	x	3.32	**3.86**	**4.91**	x	1.68
LO	2.40	2.77	2.53	x	3.55	3.04	x	2.95	**4.56**	**4.94**	3.13	1.69

Z scores of the local maxima for each contrast in each ROI. Significant values are in bold ($p<0.0001$, uncorrected). The x denotes no voxel surviving a significance threshold of $p < 0.05$, uncorrected. R= right hemisphere; L=left hemisphere.

compared to activations in areas V4, VOT and LO (Fig 2.2.5B and D, Table 2.2.1). Contrary to the other analysed areas, V1 was mainly deactivated by forms, i.e. responses were stronger for condition D (Fig 2.2.5B and D, Table 2.2.1). However, in spite of the general deactivation, we found subregions of V1 more responsive to forms (Fig 2.2.5B, Table 2.2.1). Therefore, for the ROI analysis, we averaged signal of those voxels with positive values for the contrast [R - D]. We first analysed the pattern of responses elicited by the conditions in Group 1, where the only variable between them is the complexity of the form (see Materials and Methods). Results of this analysis are shown in Fig 2.2.6A. A repeated measures ANOVA revealed a significant main effect of Stimulus Type (L1, A1, R) [$F_{(2,10)}=28.55$, $P<0.0001$] and ROI [(V1, V2, V3, V3A, V4, LO and VOT), $F_{(2.1,10.6)}=4.48$, $P<0.05$] and no significant interaction Stimulus Type*ROI [$F_{(2.9,14.6)}=2.645$, $P=0.089$]. The

lack of an interaction between Stimulus Type and ROI argues against different levels of forms being processed in different manner in each particular visual area. Contrasts analysis showed significantly stronger responses for condition R than condition L1 in all ROIs (Table 2.2.2).

Table 2.2.2. Post-hoc contrasts results of the ROI analysis.

	R vs L1		R vs A1	
	$F_{(1,5)}$	P	$F_{(1,5)}$	P
V1	**8.43**	**<.01**	**14.93**	**<0.05**
V2	**59.2**	**<0.001**	**28.15**	**<0.005**
V3	**117.4**	**<0.0001**	**95.75**	**<0.0001**
V3A	**120.56**	**<0.0001**	**32.92**	**<0.005**
V4	**22.17**	**<0.005**	**20.2**	**<0.01**
LO	**30.12**	**<0.005**	<1	>0.5
VOT	**13.49**	**<0.05**	1.33	0.30

Significant results are shown in bold.

Significantly stronger responses for condition R than condition A1 were also observed in V1, V2, V3, V3A and V4, but not in LO and VOT. Nevertheless, the contrast [A1 - L1] was significant only in area V3 ($F_{(1,5)}$=7.83, P<0.05). Overall, this analysis shows a general trend of responses rhomboids > angles > lines for all areas. It should be noticed that the same trend is observed if the ROI analysis is performed extracting the times series of all voxels significantly active with any visual stimuli (F-test for all twelve conditions), as shown in Fig 2.2.6B. The difference in this case is that activations evoked by any form condition not always exceed those evoked by the dots condition D (shown as baseline level). In this case, a repeated measures ANOVA on the responses across all ROIs showed significant effects for Stimulus Type ($F_{(2,10)}$=17.97, p<0.001), ROI ($F_{(6,30)}$=6.77, p < 0.001) and a significant interaction between Stimulus Type and ROI ($F_{(12,60)}$=2.04, p <0.05), reflecting the fact that differences between conditions are not significant in areas V1 and V2. It should be noticed that even when the trend of activation rhomboids > angles > lines is still

preserved, a different criteria of voxel selection changes the significance of the results.

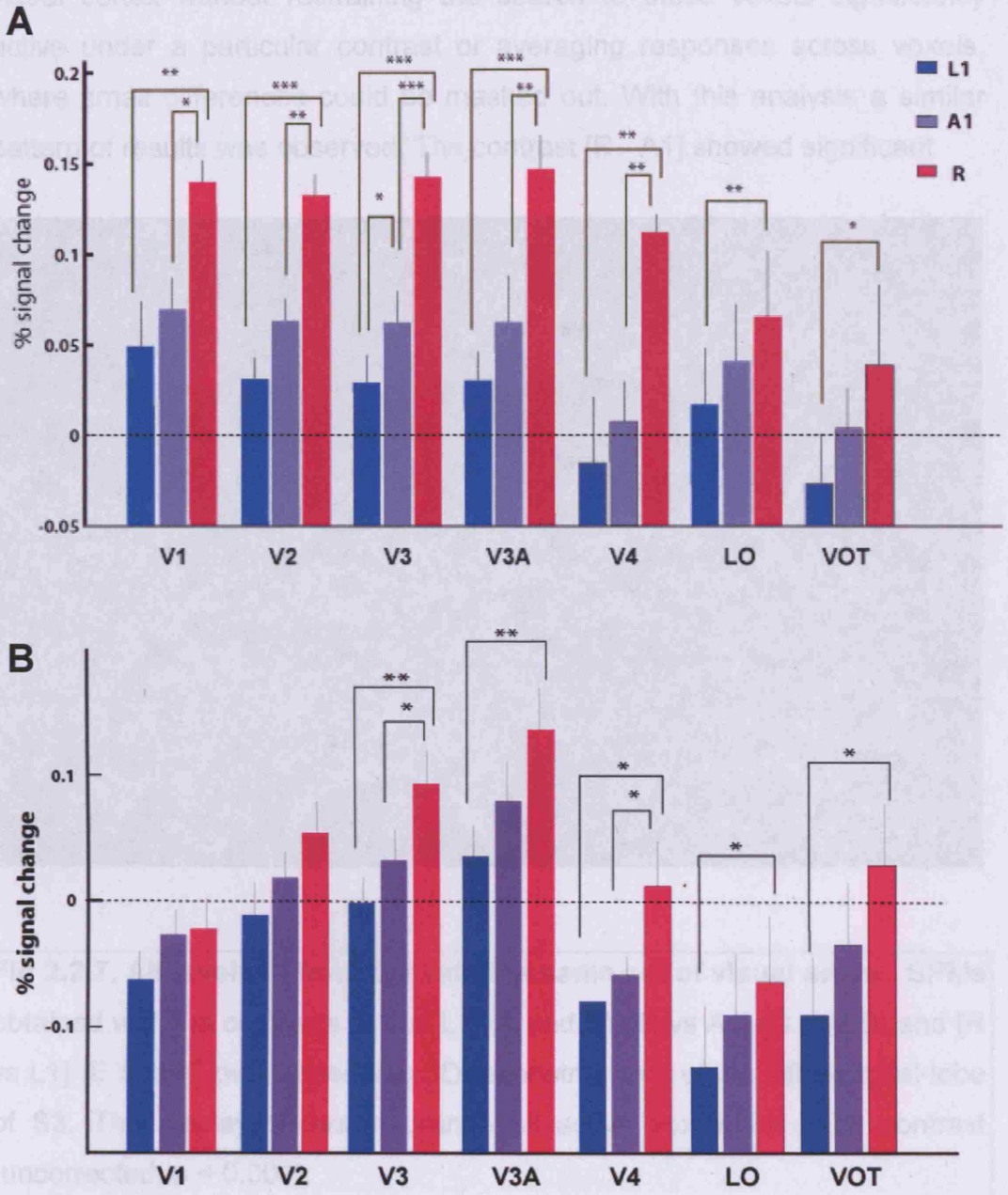


Fig 2.2.6. fMRI responses for conditions L1, A1 and R in retinotopic visual areas, LO and VOT. Bars show normalised average percent signal change across sessions (8) and subjects (6) for each condition. For each ROI, the % signal change was obtained from voxels with values above $p<0.05$ for the contrast [RvsD] **(A)** or $p < 0.001$ for an F-contrast for all 12 conditions **(B)**. *p<0.05; ** p<0.01; *** p<0.001.

This is one of the main reasons that encouraged us to also perform a single-subject analysis, where we look for differences between conditions in all the visual cortex without restraining the search to those voxels significantly active under a particular contrast or averaging responses across voxels, where small differences could be masked out. With this analysis a similar pattern of results was observed. The contrast [R - A1] showed significant

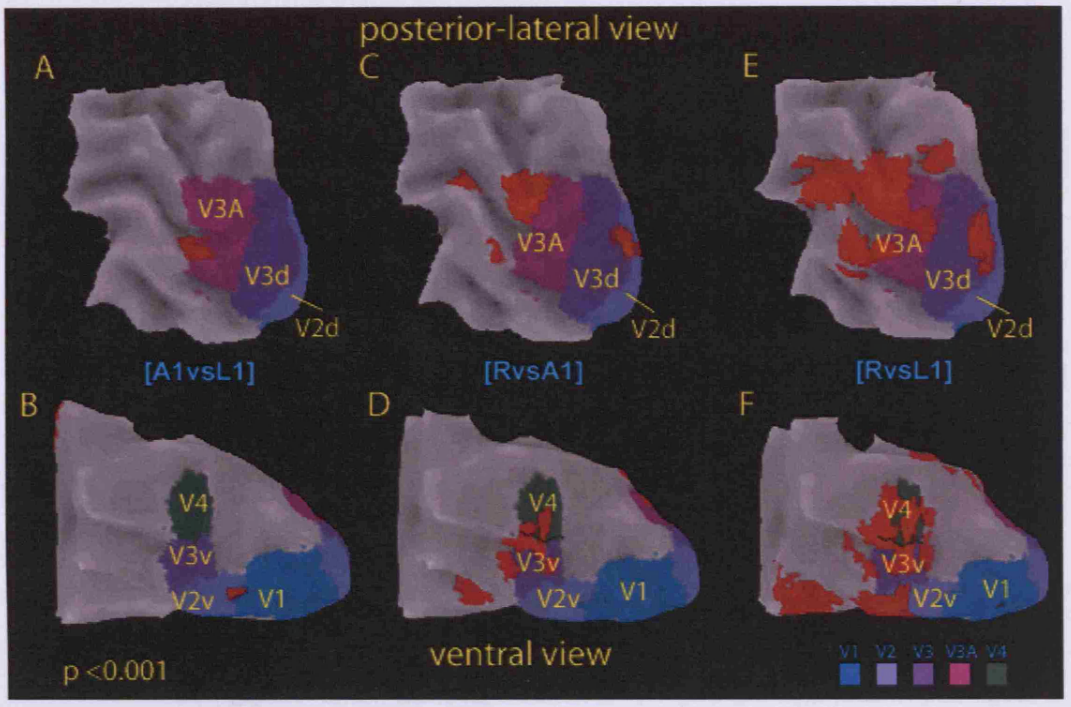


Fig 2.2.7. All levels of form activate the same set of visual areas. SPMs obtained with the contrasts [A1 vs L1] (A and B), [R vs A1] (C and D) and [R vs L1] (E and F) overlapped into 3D-reconstructions of the left occipital lobe of S3. The display shows in orange all active voxels for each contrast (uncorrected, $p < 0.001$).

activations in areas V3A and V3, and less consistently across subjects in V2 and V4 (Fig 2.2.7C and D; Table 2.2.3). With the contrast [R - L1], we observed the same spatial pattern of activations, but with more voxels reaching the significance threshold (Fig 2.2.7E and F; Table 2.2.3). With the contrast [A1 - L1] activations were mainly observed in areas V3A, V1 and V2 (Fig 2.2.7A and B, Table 2.2.3), but, in most of the subjects, these were not strong enough to reach the significance threshold (uncorrected, p <

0.0001). Parameter estimates for the maxima of the contrast [R - L1] for all subjects are shown in Fig 2.2.8.

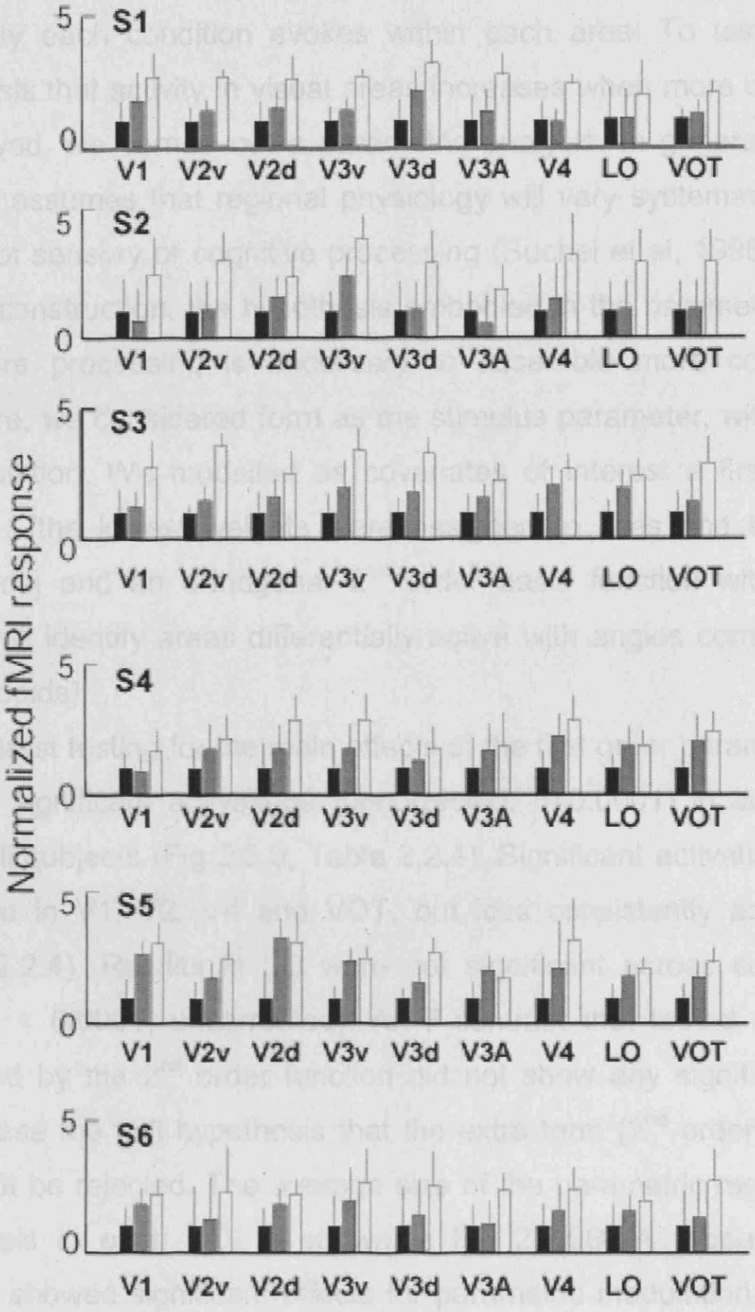


Fig 2.2.8. Parameter estimates from the maxima for the contrast [R vs L1] in areas of the left hemisphere. Bars represent % global signal change (GSC) ± s.e.m. normalised to the value obtained in condition L1. The GSC represents the intensity change in a voxel scaled to the global mean intensity (GMI), i.e. the mean of all voxels of all images in a session.

Our results show a general trend of activation rhomboids > angles > lines within each area. They suggest that lines, angles and rhomboids are processed by all visual areas in parallel, the difference being in the amount of activity each condition evokes within each area. To test formally the hypothesis that activity in visual areas increases when more complex forms are viewed, we carried out a parametric analysis. In general, this type of analysis assumes that regional physiology will vary systematically with the degree of sensory or cognitive processing (Buchel et al, 1998). In the case of form construction, the hypothesis embodied in the parametric analysis is that more processing is necessary to assemble more complex forms. Therefore, we considered form as the stimulus parameter, with three levels of modulation. We modelled as covariates of interest a first order linear function [the lowest weights were assigned to lines and the highest to rhomboids] and an orthogonal 2^{nd} order basis function with a U–shape (which will identify areas differentially active with angles compared to lines or rhomboids).

A T-contrast testing for the main effects of the first order parameter revealed bilateral significant activations (uncorrected, $p<0.0001$) in areas V3A and V3, in all subjects (Fig 2.2.9, Table 2.2.4). Significant activations were also observed in V1, V2, V4 and VOT, but less consistently across subjects (Table 2.2.4). Results in LO were not significant across subjects (Table 2.2.4; $p < 0.0001$, uncorrected). An F-contrast that tested for activations explained by the 2^{nd} order function did not show any significant result, in which case the null hypothesis that the extra term (2^{nd} order) is redundant could not be rejected. The average size of the parametric regressors for all the voxels in each ROI is shown in Fig 2.2.9B. A repeated measures ANOVA showed significant effects for parametric modulation (1^{st} order, 2^{nd} order) [$F_{(1,5)}=19.22$, $P=0.007$] and ROI [$F_{(4.7,23.4)}=3.65$, $P=0.015$], and no significant interaction [$F_{(6,30)}=1.19$, $P=0.337$]. Post hoc one-sample t-tests revealed a significant linear parametric modulation in all visual areas, but not a significant 2^{nd} order one (Table 2.2.5). These results demonstrate that there is an increase in neural activity that correlates positively with the increase in form complexity through each visual area, more significantly in areas V3A and V3.

Table 2.2.3. Z scores for the contrasts [A1 - L1], [R - A1] and [R - L1]

	[A1 - L1]						[R - A1]						[R - L1]					
	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**
V1	**4.22**	x	3.6	3.19	**4.35**	1.98	**4.37**	3.14	2.90	1.90	2.13	2.95	**4.22**	3.10	3.38	2.92	**4.84**	2.80
V2	3.32	3.52	3.12	**3.79**	**4.87**	2.67	**4.82**	**4.60**	3.30	2.28	**3.95**	3.20	**4.27**	**4.26**	**5.08**	2.46	**6.41**	3.59
V3	2.96	3.12	2.90	2.73	**5.50**	2.47	**4.06**	**4.60**	**4.50**	3.52	**4.63**	**3.98**	**4.83**	**4.40**	**5.56**	**4.31**	**7.37**	**3.95**
V3A	2.16	3.51	3.42	2.76	**4.3**	2.03	**4.27**	**4.69**	**3.79**	3.30	1.74	**4.03**	**4.59**	**5.29**	**5.43**	**4.31**	**4.43**	**3.95**
V4	1.89	x	2.90	2.45	3.08	**3.21**	2.8	**4.76**	**4.82**	2.01	**4.65**	3.30	3.51	**4.09**	**6.33**	1.95	**6.00**	3.30
VOT	3.19	2.17	2.85	3.10	3.64	2.39	3.44	**3.88**	1.97	2.00	x	3.38	3.42	**4.40**	3.20	2.23	**4.07**	3.45
LO	2.81	1.88	3.06	2.71	3.04	1.89	2.56	3.59	3.05	1.68	2.42	3.47	**3.72**	**4.09**	**3.48**	**4.00**	**4.28**	**3.76**

Z scores of the local maxima for each contrast in each ROI. Significant values ($p<0.0001$, uncorrected) are in bold. The x denotes no voxel surviving an uncorrected threshold of $p < 0.05$.

Table 2.2.7. Z scores for the main effects and interactions of groups 1 and 2 of angles and lines.

	Group 1 – Group 2						Group 2 – Group 1						[A1-L1] - [A2-L2]						[A2-L2] - [A1-L1]					
	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**
V1	x	2.77	3.03	x	2.19	x	**6.15**	2.96	**4.35**	3.54	**4.65**	**4.04**	3.42	1.70	3.09	2.25	**4.68**	1.92	2.4	2.75	1.32	2.05	x	3.25
V2	x	3.25	2.04	2.13	2.77	x	**5.58**	x	3.01	**4.04**	**4.65**	**4.32**	3.36	3.24	3.12	2.48	**3.74**	2.08	2.56	2.75	1.86	1.54	x	3.55
V3	1.79	**3.82**	3.50	2.67	3.29	2.34	**4.44**	x	2.09	2.21	2.98	**3.74**	3.15	2.42	2.63	X	3.50	x	2.84	1.84	2.40	3.50	2.18	**4.02**
V3A	x	**6.55**	**3.89**	2.60	**4.28**	3.60	**3.88**	x	x	1.89	x	3.20	2.93	3.10	2.29	2.26	3.58	1.76	3.27	1.89	2.40	1.68	x	**4.01**
V4	x	2.85	2.10	2.14	2.15	x	**4.76**	x	1.97	x	2.45	**3.75**	3.10	x	2.29	2.26	3.01	1.76	2.43	2.45	x	3.09	2.12	2.40
VOT	x	2.74	3.10	2.61	1.91	2.90	**4.53**	x	2.03	2.63	2.52	2.77	2.93	x	1.99	2.20	2.42	1.81	2.06	2.53	x	2.87	2.20	3.46
LO	x	3.54	**4.55**	x	3.27	2.70	2.94	x	2.37	1.83	2.22	2.25	2.69	1.88	2.52	2.64	3.32	2.31	1.71	3.22	1.80	1.92	2.50	3.09

Table 2.2.4. Z scores for the linear parametric analysis at the single-subject level.

	Positive correlation						Negative correlation					
	S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6
V1	**4.09**	3.06	3.60	2.84	**4.79**	2.75	1.84	2.74	1.68	2.16	3.24	**4.45**
V2	**4.16**	**4.43**	**5.01**	2.52	**6.43**	3.64	x	2.32	x	1.88	2.04	**5.24**
V3	**4.77**	**4.38**	**5.48**	**4.28**	**7.47**	**3.93**	x	2.07	x	1.79	2.19	3.48
V3A	**4.37**	**5.33**	**5.48**	**4.28**	**4.47**	**3.74**	2.19	x	2.76	2.39	1.65	1.81
V4	3.30	**4.15**	**6.27**	2.05	**6.19**	3.36	x	3.17	x	x	x	2.75
VOT	3.69	**4.47**	3.58	**4.18**	**4.49**	3.54	x	1.82	x	2.49	2.86	2.42
LO	3.02	2.08	3.16	2.30	**4.39**	3.38	x	3.27	2.21	2.84	x	1.97

Z scores of the maxima for each contrast in each ROI. Significant values ($p<0.0001$) are in bold. The x denotes no voxel surviving a significance threshold of $p < 0.05$.

The parametric model also allows us to test the hypothesis that activity diminishes with the complexity of form by performing the converse contrast. This analysis did not reveal, consistently, any significantly active voxel in any visual area (Table 2.2.4). As well, when we performed the contrasts [L1 - R], [L1 - A1] and [A1 - R], in order to identify areas more responsive to less complex forms than to more complex ones, no significant results were obtained (Table 2.2.6).

It is well known that in the macaque brain, the RF size of visual areas increases from posterior to anterior (Zeki, 1978c). This has also been demonstrated in humans with fMRI experiments (Smith et al, 2001).

Table 2.2.5. Post-hoc contrasts of the parametric analysis for each ROI.

	1st order		2nd order	
	$T_{(5)}$	P	$T_{(5)}$	P
V1	4.23	<0.01	1.93	0.11
V2	11.612	<.0001	1.77	0.135
V3	14.233	<.0001	1.97	0.106
V3A	15.456	<.0001	1.73	0.144
V4	5.561	<0.005	1.81	0.13
LO	6.991	<0.001	-0.137	0.896
VOT	4.273	<0.01	0.694	0.519

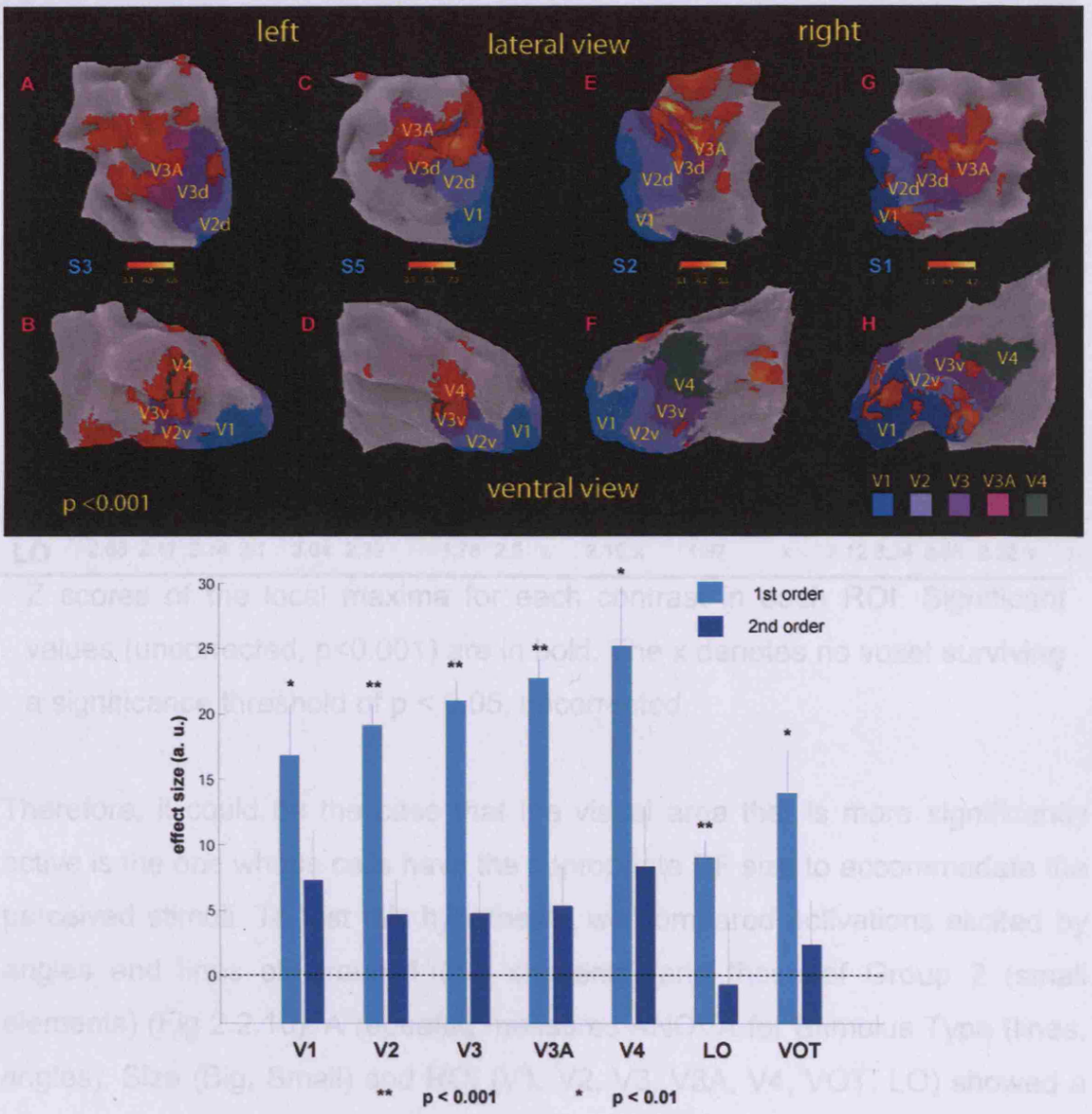


Fig 2.2.9. Parametric Analysis. Top. SPMs obtained with a T-contrast for the 1st order linear function of the parametric analysis (see Materials and Methods) overlaid on 3D reconstructions of the cortical surface of the left occipital lobes of S3 and S5 and right occipital lobes of S1 and S2. The displays include all significant voxels at uncorrected values, $p < 0.001$. **Bottom.** Average size of the contrasts testing for a 1st (cyan) and 2nd order (blue) parametric effect in each ROI. Note that the voxels included in this analysis are the same than those shown in Fig 2.2.6A. A significant effect means that a model where activity increases as function of form complexity is better than a null model where each form evokes the same amount of activity.

Table 2.2.6. Z scores for the contrasts [L1 - R], [L1 - A1] and [A1 - R]

	[L1 - R]						[L1 - A1]						[A1 - R]					
	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**
V1	x	2.76	x	2.10	3.47	**4.14**	3.34	2.99	2.29	2.03	x	2.45	2.82	3.18	2.06	2.68	**3.88**	2.93
V2	1.78	2.76	x	1.92	2.18	**4.97**	3.68	3.02	1.75	1.96	2.23	2.61	x	2.93	1.74	2.68	**3.79**	2.82
V3	x	2.28	x	1.81	2.23	3.35	3.52	2.18	x	2.11	x	2.24	x	x	x	2.59	**3.81**	2.40
V3A	2.38	x	3.02	2.38	x	2.06	3.30	x	2.46	x	x	2.05	x	x	1.65	x	2.28	x
V4	x	3.34	x	x	x	2.74	2.86	2.50	x	2.40	2.07	1.88	x	2.89	x	1.90	2.38	2.29
VOT	2.66	1.87	3.05	2.71	2.89	2.61	1.79	2.18	2.37	1.77	2.31	2.99	1.75	2.09	x	2.79	3.20	x
LO	2.68	2.17	2.14	3.1	3.64	2.39	1.78	2.5	x	2.19	x	1.97	x	3.12	3.24	2.83	2.32	x

Z scores of the local maxima for each contrast in each ROI. Significant values (uncorrected, $p<0.001$) are in bold. The x denotes no voxel surviving a significance threshold of $p < 0.05$, uncorrected.

Therefore, it could be the case that the visual area that is more significantly active is the one whose cells have the appropriate RF size to accommodate the perceived stimuli. To test this hypothesis, we compared activations elicited by angles and lines of Group 1 (big elements) and those of Group 2 (small elements) (Fig 2.2.10). A repeated measures ANOVA for Stimulus Type (lines, angles), Size (Big, Small) and ROI (V1, V2, V3, V3A, V4, VOT, LO) showed a significant main effect of Stimulus-Type ($F_{(1,5)}=6.88$, $P=0.047$) and ROI ($F_{(6,30)}=5.036$, $P=0.0001$), and a significant interaction Size*ROI ($F_{(30,6)}=5.423$, $P<0.05$). Contrasts revealed a significant interaction Size*ROI between areas V1 and V3 ($F_{(1,5)}=6.71$, $P<0.05$), V1 and V3A ($F_{(1,5)}=13.86$, $P<0.05$), V1 and LO ($F_{(1,5)}=9.32$, $P<0.05$), V2 and V3A ($F_{(1,5)}=15.81$, $P<0.05$), V2 and LO ($F_{(1,5)}=12.82$, $P<0.05$), V2 and VOT ($F_{(1,5)}=7.42$, $P<0.05$), and LO and VOT ($F_{(1,5)}=10.79$, $P<0.05$), showing that responses to forms of different sizes are different in each ROI. The absence of a significant interaction Stimulus Type*Size demonstrates that responses are qualitatively the same between each level of form complexity independently of the size of the constitutive elements.

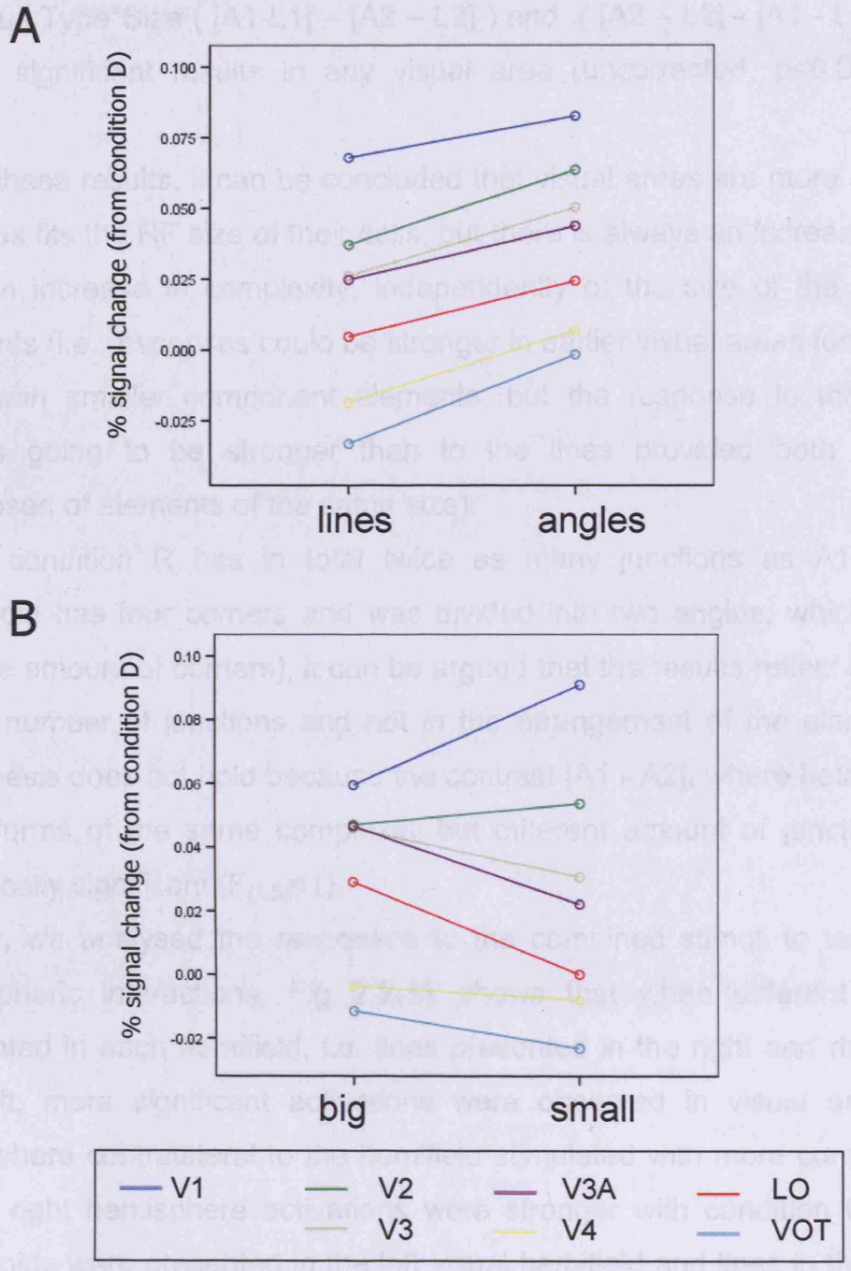


Fig 2.2.10. Effect of the size of the components of a form on the fMRI response. The figure shows, for each ROI, responses to lines (L1 and L2) or angles (A1 and A2) with elements of all sizes (**A**) and to forms with big (L1 and A1) or small (L2 and A2) components (**B**). Responses were averaged across subjects and sessions.

The single subject analysis also showed that images of Group 1 elicited more significant activations in areas with larger RFs (V3 and V3A), and those of

Group 2 in areas with smaller RFs (V1 and V2) (Table 2.2.7). The interactions Stimulus Type*Size ([A1-L1] – [A2 – L2]) and ([A2 – L2] - [A1 - L1]), did not reveal significant results in any visual area (uncorrected, $p<0.0001$; Table 2.2.7).

From these results, it can be concluded that visual areas are more active if the stimulus fits the RF size of their cells, but there is always an increase in activity with an increase in complexity, independently of the size of the constitutive elements (i.e. responses could be stronger in earlier visual areas for angles and lines with smaller component elements, but the response to the angles is always going to be stronger than to the lines provided both stimuli are composed of elements of the same size).

Since condition R has in total twice as many junctions as A1 (i.e. each rhomboid has four corners and was divided into two angles, which results in half the amount of corners), it can be argued that the results reflect an increase in the number of junctions and not in the arrangement of the elements. This hypothesis does not hold because the contrast [A1 - A2], where both conditions have forms of the same complexity but different amount of junctions, is not statistically significant ($F_{(1,5)}<1$).

Finally, we analysed the responses to the combined stimuli to test for inter-hemispheric interactions. Fig 2.2.11 shows that when different forms are presented in each hemifield, i.e. lines presented in the right and rhomboids in the left, more significant activations were observed in visual areas of the hemisphere contralateral to the hemifield stimulated with more complex forms. In the right hemisphere activations were stronger with condition RILr, where rhomboids were presented in the left visual hemifield and lines in the right one. The opposite was observed in the left hemisphere, where stronger activations were observed with the condition LIRr (Fig 2.2.11). The same trend is observed when comparing combined conditions containing angles (data not shown). To test for an interaction effect, percent signal changes were computed separately for each right and left ROI when forms were presented in the contralateral hemifield (Fcontra) in the presence of different ones in the ipsilateral hemifield (Fipsi) (Fig 2.2.12).

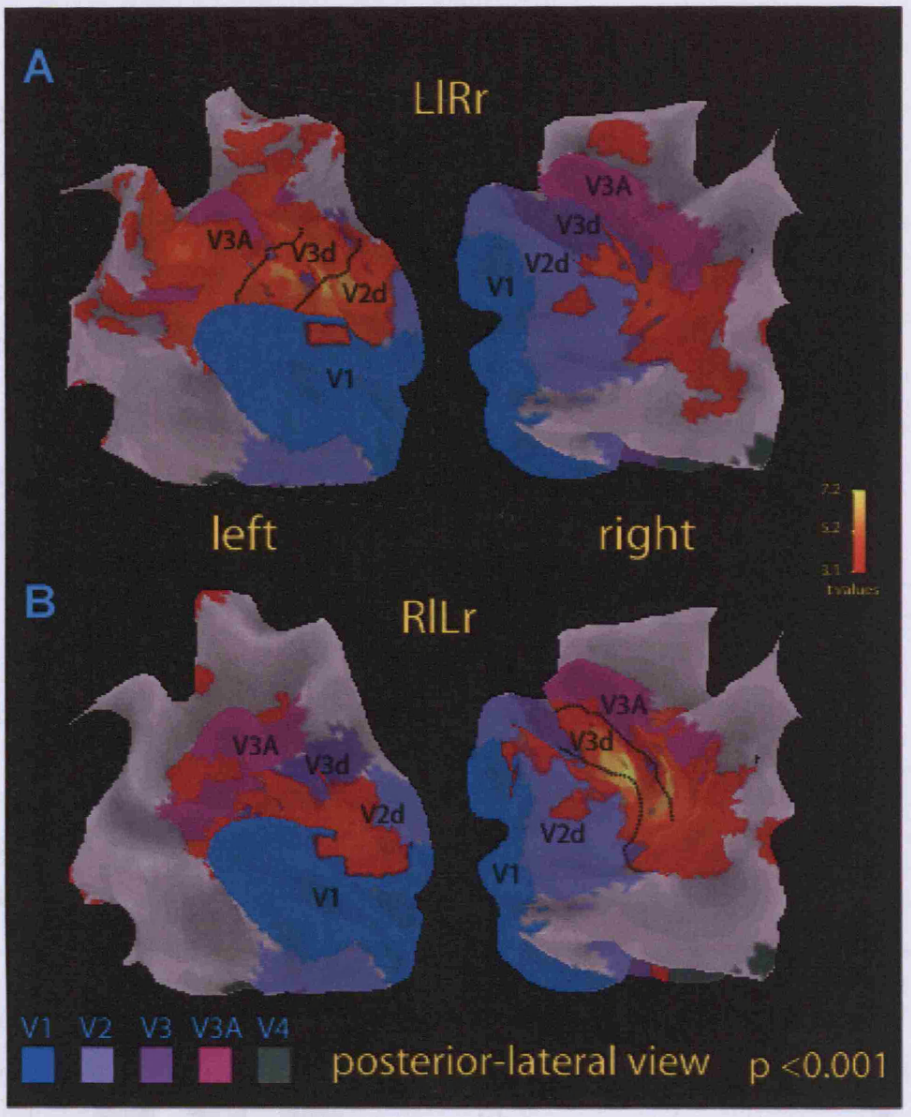


FIG 2.2.11. Activations with combined stimuli. A-B. SPMs for the contrast [LIRr vs D] and [RILr vs D] are overlapped on 3D reconstructions of the cortical surface of the right and left occipital lobes of S2. The figure shows all significant voxels at uncorrected values p<0.001.

A repeated measures ANOVA showed a significant main effect of Fcontra (L1, A1, R) [$F_{(2,10)}$=36.32, P< 0.001], Fipsi (L1, A1, R) [$F_{(2,10)}$=9.07, P< 0.01], and ROI (V1, V2, V3, V3A, LO and VOT) [$F_{(6,30)}$=2.96, P< 0.05] and no significant main effect of hemisphere (right, left) [$F_{(1,5)}$=2.19, P=0.2]. We also found a significant interaction between Fcontra * ROI [$F_{(12,60)}$=4.21, P<0.001] and a significant interaction between Fcontra * Fipsi [$F_{(4,20)}$=4.95, P<0.01].

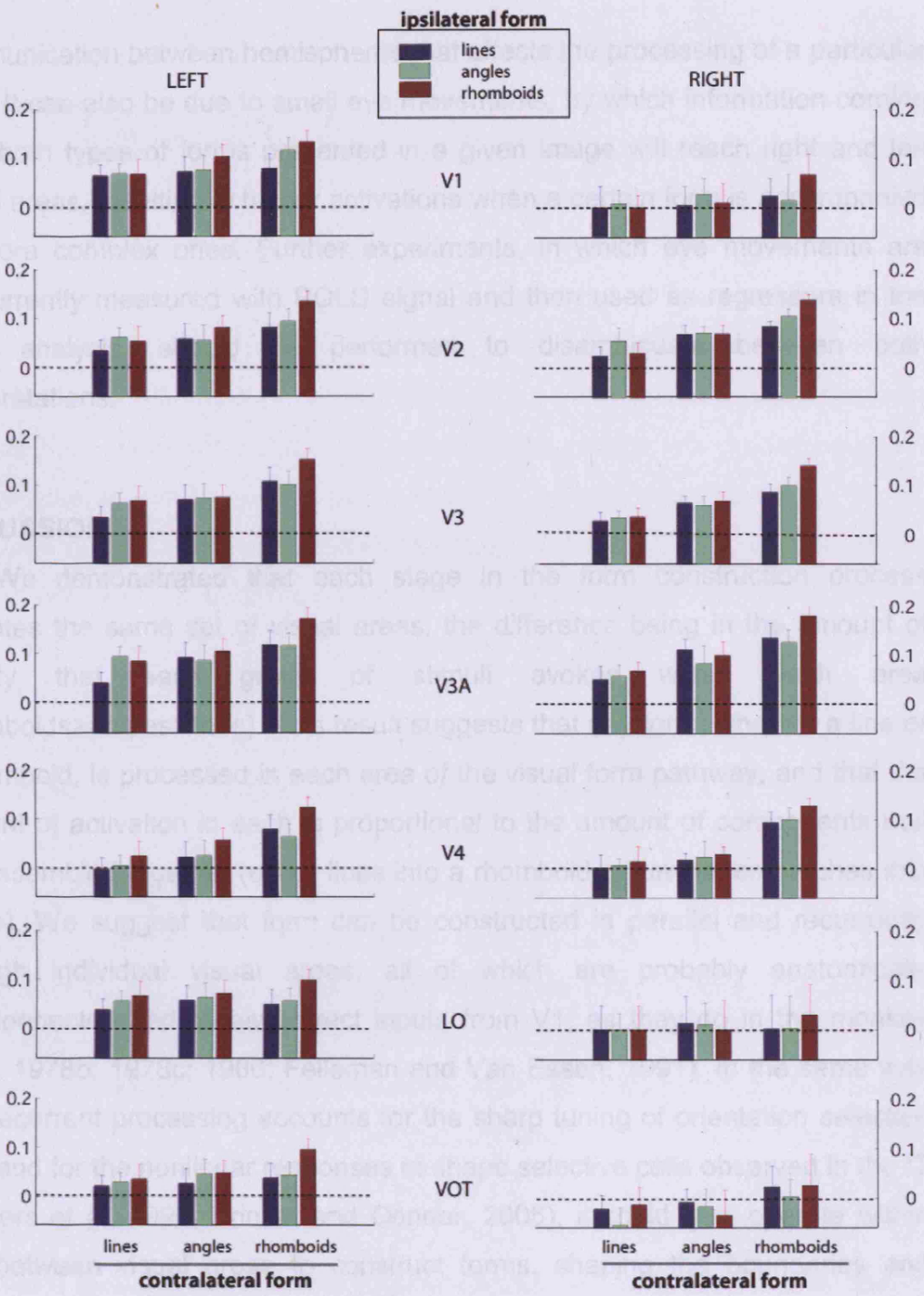


Fig 2.2.12. Inter-hemispheric interactions in form construction. The figure shows the response in visual areas of the left and right hemisphere evoked by forms presented in the contralateral hemifield in the presence of forms of different level of complexity in the ipsilateral and the contralateral hemifield.

This could be due to visual areas in one hemisphere influencing what is being processed in the opposite one, from which it can be concluded that there is a

communication between hemispheres that affects the processing of a particular form. It can also be due to small eye movements, by which information coming from both types of forms presented in a given image will reach right and left visual areas, resulting in higher activations when a certain form is accompanied by more complex ones. Further experiments, in which eye movements are concurrently measured with BOLD signal and then used as regressors in the GLM analysis, should be performed to disambiguate between both interpretations.

DISCUSSION

We demonstrated that each stage in the form construction process activates the same set of visual areas, the difference being in the amount of activity that each group of stimuli evokes within each area (rhomboids>angles>lines). This result suggests that any form, whether a line or a rhomboid, is processed in each area of the visual form pathway, and that the amount of activation in each is proportional to the amount of components that are assembled together (either lines into a rhomboid or luminance patches into a line). We suggest that form can be constructed in parallel and recurrently through individual visual areas, all of which are probably anatomically interconnected and receive direct inputs from V1, as they do in the monkey (Zeki, 1978b; 1978c; 1980; Felleman and Van Essen, 1991). In the same way that recurrent processing accounts for the sharp tuning of orientation selective cells and for the nonlinear responses of shape selective cells observed in the IT (Somers et al, 1995; Brincat and Connor, 2006), it could also operate within and between visual areas to construct forms, shaping the boundaries and domains of each one. Some neurons in IT exhibit enhanced nonlinear responses to the combined presence of two or more contour parts, consistent with a recurrent network that compares parts signals across neurons and generates inferences about multipart configurations (Brincat and Connor, 2006). Our results support this idea, where nonlinear responses and a recurrent process will both result in more neural activity for stimuli where elements are arranged into multiparts, like oriented lines arranged to form rhomboids. The response of a particular visual area is also influenced by the information

processed in the others, this is more obvious in the case of the interaction observed with ipsilateral and contralateral forms, where retinotopic responses are enhanced if more complex forms are presented to other parts of the visual field. However, this interaction could also be the result of small eye movements, a possibility only further experiments could disambiguate.

We nevertheless recognize a hierarchical organization in the visual system, since simple forms seem to be activating more strongly intermediate visual areas like V3 and V3A, than earlier or later ones. What it is not supported here is a *discrete* hierarchical process, where earlier areas will process simpler components, and once solved, this information is sent to further areas. If form construction occurs through discrete modular hierarchical process, we would have expected to see a switch in the most active area depending on the complexity of the form. Instead, we observed an increase in activity with the complexity of the form through all the tested areas (parametric analysis). The extent of the elicited activations was different between visual areas, but nonetheless present, which suggests that the construction of forms occur continuously through visual areas, as it seems to be the general mechanism of processing of object information (Humphreys et al, 1999).

Previous studies have shown activity related to global shape perception in both, 'early' and higher visual areas, suggesting that shape perception involves multiple visual areas each of which may integrate local elements to global shapes at different spatial scales (Altmann et al, 2003; Kourtzi et al, 2003b; Ban et al, 2006). Here we decomposed the construction of a form into several stages, showing how an increase in the activity in visual areas relates to more complex construction processes from early to higher visual areas, but particularly in intermediate ones. The hypothesis that form construction occurs mainly in intermediate visual areas is consistent with the behaviour observed in patients with different subtypes of apperceptive agnosia. Humphreys and Riddoch (1987) have studied for several years patient HJA, who has integrative agnosia due to a ventral extrastriate lesion. This subject cannot recognize objects and has difficulties in everyday life, but can identify simple shapes and makes good drawings of the images he is presented with (if given unlimited time). Therefore HJA's ability to construct a form seems preserved, but the mechanism to identify objects is impaired. On the other hand, SA, a patient with

form agnosia due to a dorsal extrastriate lesion, shows good naming of pictures and is much less impaired in everyday life, but fails in discriminating similar geometric shapes (Efron shape test) and makes inaccurate drawings (Riddoch et al, 2008). This shows that, with a lesion in dorsal extrastriate areas, a patient is still able to name pictures, probably using cues like colour or context, but with a ventral extrastriate lesion recognition is not possible, even when the patients are able to draw a visual scene accurately. It could then be hypothesized that an intermediate dorsal region, like V3A, performs a central role in the construction of form and perception of simple shapes, and that more anterior regions are involved in the recognition of objects, possibly using the information provided by previous retinotopic areas. Again, even when this were so, it will not be a discrete process and information will be continuously looped until solved, using both bottom-up and top-down signals.

The majority of cells in macaque early and intermediate visual areas show orientation selectivity (Hubel and Wiesel 1968; Zeki 1978c), making them the most suitable and obvious areas to construct forms defined by luminance edges, like the ones used in this study. We also found some regions in the VOT showing the same trend of activation observed in earlier visual areas. The VOT is often considered as part of the LOC, a huge complex of areas where differential responses within subregions have been obtained in several studies (Grill-Spector et al, 1999; Avidan et al, 2003; Kourtzi et al, 2003a), emphasizing the importance of making a clear definition of the activated regions. Restraining ourselves to pure anatomical definitions, VOT corresponds to the ventral surface of the occipital lobe, anterior to V4. The observed regions activated by form in the VOT could correspond to V4α, the anterior subdivision of the human colour centre, which is responsive to colour and has been shown to be coactive with areas processing faces and objects (Bartels and Zeki 2000). They could correspond as well to hemifield maps in the VO cluster, VO-1 and VO-2, also anterior to V4, that are responsive to colour and objects (Brewer, Liu et al 2005). Further studies should clarify if V4α and the hemifields in the VO cluster are actually the same area or represent functionally and anatomically distinctive regions. Our retinotopic maps did not have enough resolution to allow to identify recently described retinotopic areas LO1 and LO2 (Larsson and Heeger, 2006), but these areas are probably embedded in what we refer

here as LO, since they are located in the lateral occipital cortex dorsally to V3. Area V3B, located lateral to V3A, has been also involved in the processing of contours defined by several cues (Zeki et al, 2003) and depth (Tyler et al, 2006). We were not able to delimited areas beyond V3A, but since we delimited V3A as a complete hemifield next to V3d, it is possible that is also includes as well area V3B, as described by Press et al (2001).

A particular profile of response was found in V1, where most of the voxels seem to deactivate with form stimuli (more active for condition D than R). A simple explanation of this is that a condition with random dots (1 pixel diameter) will be a better stimulus for driving cells in V1, due to the smaller size of their RFs. We cannot totally rule out this explanation, but it will be reducing V1 responses to purely RF size-related effects, which seems unlikely to be the case since we also observe stronger responses to forms in other subregions of V1, with activity correlating positively with complexity. A previous study has shown a reduction in activity in V1 with 3D shapes stimuli compared to 2D shapes and random oriented lines (Murray et al, 2002), supporting a predictive coding theory of cortical processing. According to this theory, higher visual areas will project their predictions about the stimulus to early areas via backward connections, where they will then be subtracted from the incoming data (Rao and Ballard, 1999; Friston, 2003). If so, activity in lower visual areas should decrease when neurons in higher areas can 'explain' a visual stimulus. Our results in V1, where we observed more activation in response to random dots patterns than to forms, supports a predictive coding view. We did not observe more activation in early visual areas with lower levels of form compared to higher ones. On the contrary, parameter estimates of maxima show the same trend of activations observed in other areas (rhomboids>angles>lines). This result, at first sight, seems at odds with that of Murray et al (2002). We consider that lines, angles and rhomboids constitute well defined forms, making them suitable for an exact prediction from higher visual areas, and that a possible explanation for the discrepancy between our results and those of Murray et al (2002) could be the fact that they used forms of 4° distance from fixation, which probably could not be integrated by V1 neurons. Our stimuli comprised a wide range of sizes (0.6°-4°) making it plausible to suppose that some shapes are resolved within V1.

In conclusion, we showed that different steps in the process of form construction activate the same visual areas, the intensity of activity depending upon the complexity of the form located in a particular region of the visual field, but also influenced by those present in the rest of the scene. Our results do not support a forward hierarchical view where earlier visual areas will process simpler forms and once this information is solved, later areas will process more complex ones. Instead, they suggest that form is constructed in parallel and recurrently within and between visual areas, incorporating excitatory and inhibitory signals.

Chapter 2.3. –TOP-DOWN MODULATIONS IN THE VISUAL FORM PATHWAY.

INTRODUCTION

To learn about the environment and successfully interact with it, humans need to extract regularities from the incoming sensory information. This learning process takes place continuously, even without an individual's awareness (Reber, 1967; Saffran et al, 1996; Chun and Jiang, 1998; Perruchet and Pacton, 2006), and all levels of sensory and cognitive processing contribute to it.

Perceptual grouping rules, described by Gestalt psychologists (Wertheimer 1923; Kofka, 1935), summarize many of the regularities extracted by the visual system. One of the most striking is collinearity, since in the environment edges which are collinear will have a greater chance of being part of the same figure than those which are not. The assignment of edges is a key part of processes the brain performs automatically, like figure-ground segregation and object identification. Higher cognitive processes like categorization and recognition are also mediated through the extraction of known regularities, by creating memory traces of particular arrangements of features and their associations with semantic concepts, either for an individual familiar object (i.e. the coffee machine we use every morning), or objects of the same category, where some common properties allow us to group them together (i.e. any coffee machine).

Previous studies have shown that the more the extractable patterns of an image, the more the activity and the recruitment of further visual areas. Altmann et al, (2003) demonstrated that collinear patterns of lines cause higher activations than random ones in early visual areas and the LOC. For more complex stimuli, it has been well demonstrated that objects activate more strongly intermediate areas and the LOC as compared to scrambled images (Malach et al, 1995; Grill-Spector et al, 1998b), with real objects causing higher activations than non-real objects in the middle occipital, inferior temporal, middle temporal, fusiform and inferior frontal cortex (Price et al, 1996; Vuilleumier et al, 2002).

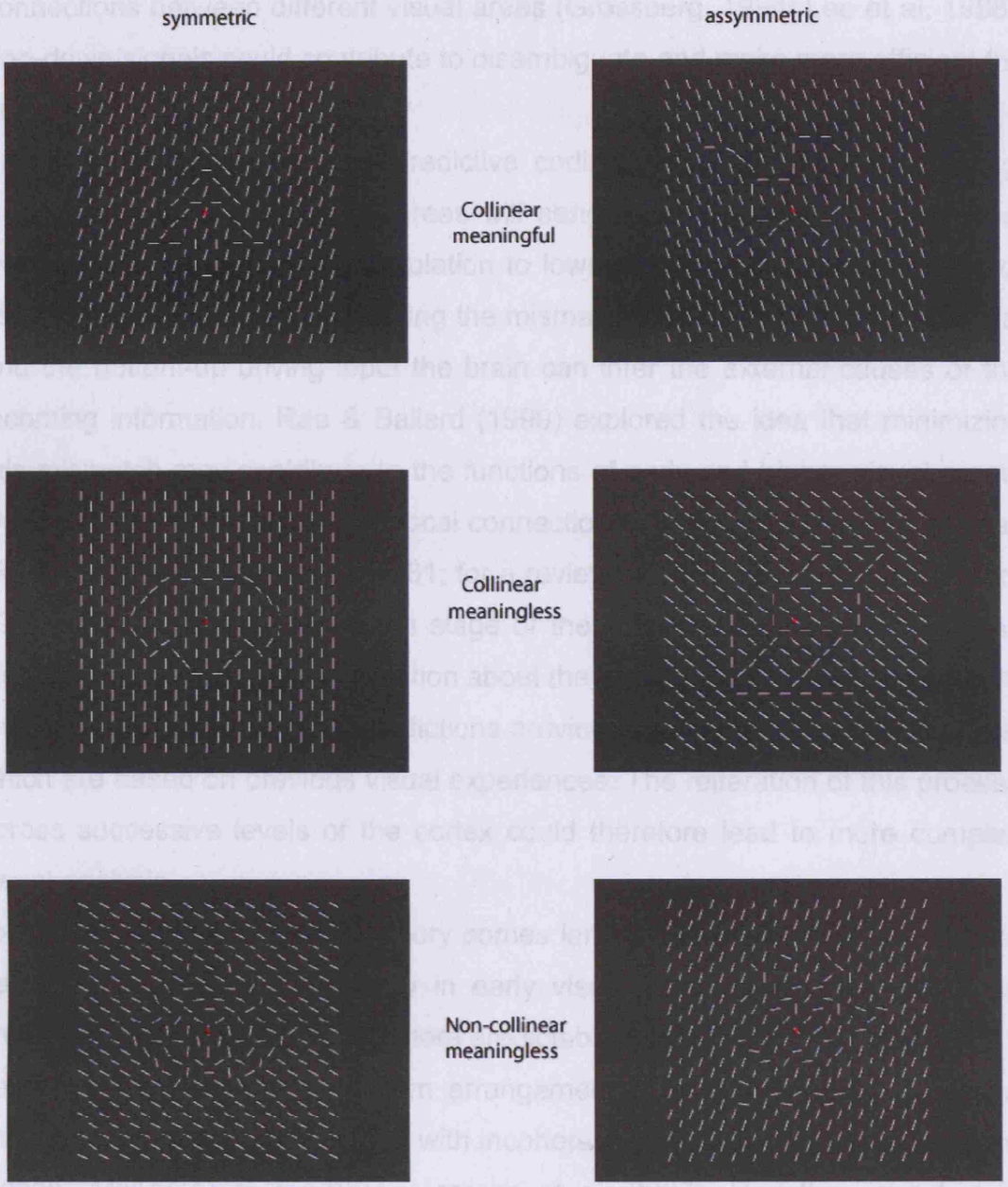


Fig 2.3.1. Stimuli. Examples of the stimuli used in the experiment. There were three conditions: collinear meaningful (MF), collinear meaningless (ML) and non-collinear meaningless (NC).

Therefore, regularities or known characteristics of an image are extracted at different levels of visual processing to ultimately determine the causes of the sensory stimulation. Stages in the visual processing of scenes, like figure-ground segregation and object recognition, are thought to interact with each other through recursive loops, mediated in the brain by forward and backward

connections between different visual areas (Grossberg, 1994; Lee et al, 1998). Top-down signals could contribute to disambiguate and make more efficient the processing in earlier areas.

This is in accordance with a predictive coding theory of neural processing, which states that higher-level areas will send signals representing the most likely cause of the sensory stimulation to lower level areas (Rao and Ballard, 1999; Friston, 2003). By minimizing the mismatch between the top-down signal and the bottom-up driving input the brain can infer the external causes of the incoming information. Rao & Ballard (1999) explored the idea that minimizing this mismatch may contribute to the functions of early and higher visual areas. On the basis of the heavy reciprocal connections between cortical visual areas (Rockland and Pandya 1979, 1981; for a review see Felleman and Van Essen, 1991), they suggested that each stage of the visual cortex may compute the discrepancy between the information about the actual visual scene coming from lower-level visual areas and predictions arriving from higher-level visual areas, which are based on previous visual experiences. The reiteration of this process across successive levels of the cortex could therefore lead to more complex visual analysis.

Experimental support for this theory comes largely from fMRI studies of visual perception, showing that activity in early visual areas is stronger with less predictable stimuli compared to more predictable ones. Higher activations in V1 have been shown with a random arrangement of lines compared to closed shapes (Murray et al, 2002), and with incoherent motion compared to coherent motion (McKeefry et al, 1997; Harrison et al, 2007). Results presented in Chapter 2.2 also support this fact, since activity in V1 was stronger with the random dots stimuli than with any of the form conditions. This is as expected if higher areas send top-down signals that will reduce the activity of earlier areas each time that it is easier to extract more recognizable patterns from a stimulus. On the other hand, single unit experiments have shown that the response rate of a neuron increases when an oriented stimulus presented within its receptive field is accompanied by a second collinear stimulus, while the same oriented stimulus presented orthogonal to the main axis will produce inhibition or at least less facilitation (Kapadia et al, 2005). Responses in V1 are also higher when the same texture is part of a figure compared to when is part of the background

(Lamme, 1995). Due to the latencies of this increase in activity, this process is thought to be mediated by excitatory top-down influences. fMRI data also shows increased activations in V1 with collinear patterns (Altmann et al, 2003), and with more complex forms compared to simple ones (Chapter 2.2). To account for these results, one need to postulate two populations: one that signals the mismatch and another that signals the possible causes of the percept (Friston, 2005). This second population also accommodates theories where signals through backward connections enhance the activity of those units that are coding relevant features of the scene, like the case of adaptive resonance (Grossberg, 1999).

Visual areas specialised in different functions can generate different kinds of top-down signals, each of which can contribute to explain a different part of the sensory stimulation (one area could determine the shape of the figure corresponds to two conical parts and further identify it as coffee machine).

With this in mind, we wanted to investigate if top-down signals: a) are generated at each stage of processing of the visual scene, b) could arrive at a particular area from more than one higher area, and c) are inhibitory and/or excitatory depending on the population of neurons they target.

In order to test these hypotheses an experiment was performed where subjects were presented with three different groups of images composed of oriented lines, each group containing a different level of recognizable patterns. In the first condition (NC, non-collinear), some of the lines were oriented differently from the background, forming a non-sense figure that could be segregated from the background lines (Fig 2.3.1, top). In the second condition (ML, meaningless), the lines that were differently oriented were collinear, giving the images an extra recognizable regularity (i.e. collinearity) even though they were still non-sense. In the third group (MF, meaningful), the lines were reoriented collinearly as well, but here they represented man-made objects. Therefore subjects could associate the figures in this group with previously learnt semantic concepts.

We expected that in the presence of some recognizable patterns, like collinearity and meaning, an interpretation of each of these feats would be generated in different visual areas. These representations could be sent via top-down mechanisms to earlier areas to explain different attributes of the

sensory stimulation, and in this case to contribute to perform successfully a symmetry judgement task. An analysis of coupling between different areas was performed to determine the generation of top-down signals and its nature.

METHODS

Stimuli and Experimental Design.

Seventeen subjects participated in the study. Results from one subject were excluded from the analysis due to problems with the recording of the button presses and two others were excluded because they fell asleep during the experiment. Therefore, the results of 14 subjects [ten male, mean age=28.5 (range 23-40), two left handed] were analysed and are reported here.

The stimuli were visual images consisting of a rectangular area (5 degrees vertical and horizontal) filled with 313 oriented white lines against a grey background. They were projected using an LCD projector onto a screen located at a distance of 60cm, which subjects viewed through an angled mirror. There were three stimuli conditions: collinear-meaningful (MF), collinear-meaningless (ML) and non-collinear-meaningless (NC); each of which was composed of seventy individual images. For the MF condition some of the lines were reoriented and made collinear to form abstract representations of meaningful objects (Fig 2.3.1). In the ML condition lines were collinear as well, but in this case they represented figures with no semantic meaning (Fig 2.3.1). Stimuli in the ML and MF groups were matched to have the same amount of straight and round components, and the same symmetry in the vertical, horizontal and diagonal axis. An analysis of the spatial frequency components showed no significant differences in the power spectrum between both groups of images ($p < 0.05$, corrected).

Stimuli of condition NC were generated by recombining the lines of all the stimuli of MF and ML into new non-collinear meaningless figures, keeping the mean change in orientation for each position constant between groups, as well as the mean amount of lines oriented differently from the background and the symmetry of the figures. On average, there were 33.31 ± 1.33 s.e.m. lines/per image oriented differently from the background in the MF group, 32.13 ± 1.26 in the ML and 32.74 ± 2.36 in the NC.

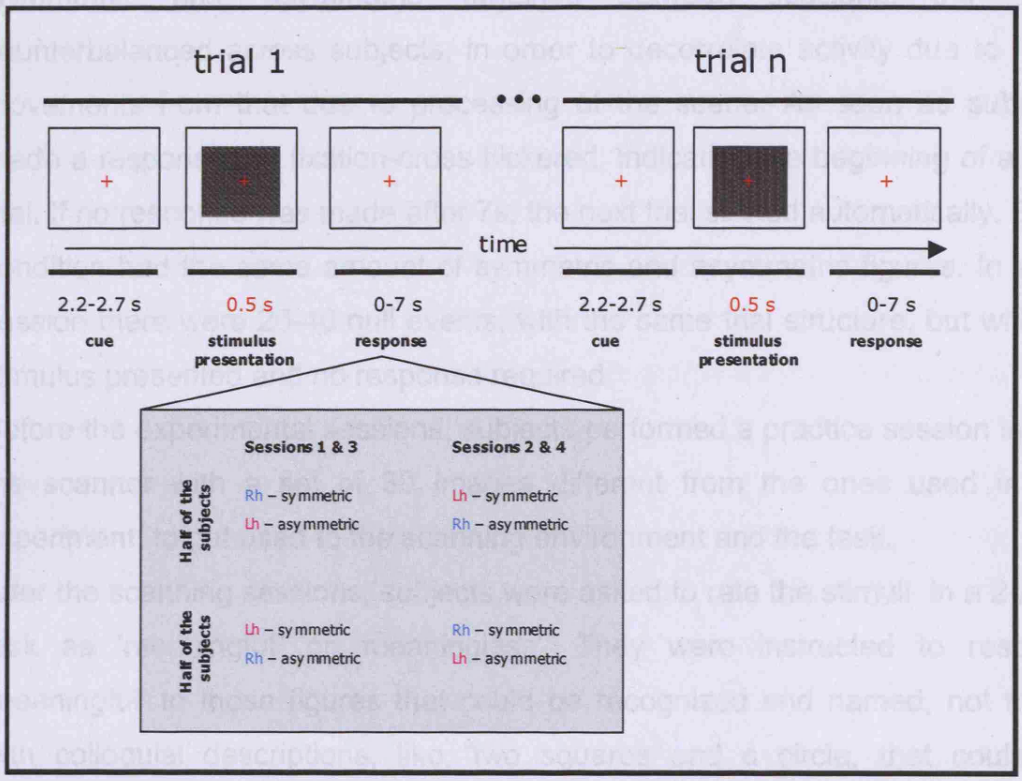


Fig 2.3.2. Schematic representation of each experimental trial. After a period of 2-2.7s, a stimulus was presented for 0.5s after which subjects had to respond 'symmetric' or 'asymmetric' with their right or left hand depending on the session. Immediately after the response the fixation cross flickered indicating the beginning of a new trial.

Subjects participated in four scanning sessions of 8-10 min each. The whole set of 210 stimuli (70 per condition), was presented twice: once in the first two sessions and for a second time in the following two. The order of the images was counterbalanced across sessions and subjects. The flicker of a red central fixation-cross indicated the beginning of a new trial (Fig 2.3.2). After a fixation period of 2-2.7s, a stimulus image was presented for 500ms. Subjects' task was to determine, in a two-alternative forced choice (2-AFC) task, if the figure in the images was symmetric or asymmetric in the vertical axis. The task was chosen to ensure subjects will pay attention to the whole image and extract the figures from the background, without making any semantic evaluation. They were instructed to answer as fast as possible, without compromising accuracy, by pressing a button with their right or left hand. The hand used to respond

'symmetric' and 'asymmetric' switched between sessions, the order counterbalanced across subjects, in order to decorrelate activity due to hand movements from that due to processing of the scene. As soon as subjects made a response the fixation-cross flickered, indicating the beginning of a new trial. If no response was made after 7s, the next trial started automatically. Each condition had the same amount of symmetric and asymmetric figures. In each session there were 20-40 null events, with the same trial structure, but with no stimulus presented and no response required.

Before the experimental sessions, subjects performed a practice session inside the scanner with a set of 30 images different from the ones used in the experiment, to get used to the scanning environment and the task.

After the scanning sessions, subjects were asked to rate the stimuli in a 2-AFC task as 'meaningful' or 'meaningless'. They were instructed to respond 'meaningful' to those figures that could be recognized and named, not those with colloquial descriptions, like "two squares and a circle, that could be something". For this task, the whole set of stimuli was presented once over two sessions, the order counterbalanced between sessions and subjects. The presentation time and interstimulus interval was the same as that for the main experiment.

Eye Movements

To demonstrate that there were no differences in eye movements between conditions, we conducted a separate experiment outside the scanner. Six different subjects took part in two experimental sessions, where the whole set of 210 stimuli (70 per condition) was presented once. The experiment had the exact same design as the one carried out in the scanner, where subjects had to respond, in a 2-AFC task, if the figures were symmetric or asymmetric. The percentage of eye movements bigger than 0.5 degrees from fixation was: a) Y-axis -- 36.2% ± 4.9% for MF, 35% ± 5.2% for ML and 35.3% ± 4.4% for NC; b) X-axis -- 15.5% ± 6.1% for MF, 18.8% ± 8.7% for ML and 14.2% ± 5.8% for NC; with no significant differences between conditions [$F_{(2,10)}<1$ and $F_{(2,10)}=1.79$, $P=0.21$, respectively]. A separate repeated-measures ANOVA revealed no significant differences in grand mean eye position from fixation between conditions ($F_{(11,55)}<1$, $P=0.65$).

Imaging

Functional and anatomical images were acquired as described in Chapter 2.1. Images were preprocessed and analysed using SPM2 and SPM5 software (**http://www.fil.ion.ucl.ac.uk/spm**). They were realigned to the first volume of the first experimental session, resliced to a final voxel resolution of 3x3x3mm, normalised to the normalised MNI reference brain in Tailarach space and spatially smoothed with a Gaussian kernel of 8mm full width at half maximum.

Main Analysis

The experiment was designed in an event-related manner with a total of 3 different conditions: collinear-meaningful (MF), collinear-meaningless (ML) and non-collinear-meaningless (NC). For every condition, each stimulus presentation was modelled as a stick-function (a boxcar of duration 1/16th of the TR), then convolved with SPM2's canonical haemodynamic response function (HRF) and entered into a multiple regression analysis to generate parameter estimates for each regressor at every voxel. The natural logarithm of the reaction times for each stimulus presentation was included as a first order modulator of no interest for each experimental condition, to exclude activations due to differences in the difficulty of the task between conditions (see Results). Null events were modelled in a separate regressor, with the onset being specified at the time the fixation period finished. The head movement parameters obtained during the realignment step and the subjects' button presses were included in the analysis as events of no interest. Data were high-pass filtered with a low-frequency cut-off of 1/128 Hz to remove low-frequency signal drifts. Contrast images were created for each subject and entered separately into voxel-wise one sample t-tests or one-way ANOVAs, depending on the case, in a random effects analysis (RFX) (Penny et al, 2003), each constituting a group level SPM.

For display purposes, SPMs are shown and summarized in tables at a threshold of $p < 0.001$, uncorrected for multiple comparisons, but activations are discussed if they survive whole brain correction for multiple comparisons at $p < 0.05$ or a small volume correction (SVC) if we had a strong anatomical hypothesis. The SVC procedure, as implemented in SPM5 using the family-

wise error (FWE) correction ($p < 0.05$), allows results to be corrected for multiple non-independent comparisons with a defined region of interest.

Dynamic Causal Modelling

To test for changes in connectivity between brain areas under the different experimental conditions we performed an effective connectivity analysis using Dynamic Causal Modelling (Friston et al, 2003; Appendix III), as implemented in SPM5. As explained in the Appendix III, given a certain architecture of brain areas and connections, DCM models the activity at the neuronal level and transforms it into area-specific BOLD signals using a haemodynamic model of fMRI measurements. Coupling parameters between these areas, under the experimental conditions, are then estimated using a Bayesian estimation scheme, such that the BOLD signals obtained with the joint forward model are as similar as possible to the observed BOLD responses (Friston et al, 2000; Mechelli et al, 2001).

Three kinds of parameters are estimated in DCM: 1) direct, extrinsic inputs to the system (i.e. the effect of all visual stimulation); 2) "intrinsic" or "fixed" connections that couples the states between the regions (i.e. the connectivity strength between area *x* and area *y*); 3) modulatory parameters that model the changes induced by the experimental manipulations in the connections between regions (i.e. the additive change a certain manipulation, like collinearity, is having in the strength of the connection between area *x* and area *y*). All parameters correspond to rate constants of the modelled neurophysiological processes with units in Hz (see Appendix III).

Time series were extracted from four regions of interest in the left hemisphere (see Results): medial middle occipital (mMO), lateral middle occipital (lMO), posterior inferior temporal (pIT) and anterior inferior temporal (aIT). Subject specific timeseries for each ROI were extracted as the principal eigenvariate of the responses across all sessions. For each subject, each ROI was defined based on anatomical and functional basis. First, the centre of each ROI was determined as the maxima in that subject for the particular contrast of interest, with the stipulation that it should not be further away than 25mm from the group maxima for that particular contrast and ROI. Time series were then extracted from voxels located in 6-mm radius sphere centred on the ROI maxima for each

subject, with exception of one subject were a 3mm radius sphere was used for areas mMO and lMO, in order to avoid overlap between the voxels of each area. The maxima for mMO was identified using the contrast [ML - MF], for lMO and pIT using the contrast [ML - NC] and for aIT with the contrast [MF - ML]. We wanted to extract only time series of the most significantly responsive voxels of each ROI. For that reason, we used as a cut-off threshold the lowest p value that resulted in at least five voxels significantly active in an F test of all the conditions of interest. Table 2.3.1 lists the thresholds used for each ROI in all subjects.

Table 2.3.1. Cut-off thresholds for ROI definition.

	ROI			
P<	mMO	lMO	pIT	aIT
0.001	10	10	14	13
0.002	1	0	0	1
0.005	1	1	0	1
0.01	1	0	0	1
0.05	1	3	0	1

The table shows the number of subjects for which each P value was used as a threshold to define the ROIs for the DCM analysis. Medial middle occipital (mMO); lateral middle occipital (lMO); posterior inferior temporal (pIT); anterior inferior temporal (aIT).

The DCMs were designed to be composed of the specified four areas, fully and reciprocally connected. We modelled as driving inputs all the visual presentations (conditions NC, ML and MF), entering the system through areas mMO and lMO. There were two modulatory influences: collinearity, which included conditions ML and MF, and meaning, which included only condition MF. These modulatory effects were allowed to have an effect on any connection, in order to let the model to determine which of these effects was significant.

Parameters for the inputs, fixed connections and modulations, were obtained for each subject and used to make statistical inferences at the group level with a one sample t-test. DCM estimation was performed using SPM5 software. Group analysis was performed using MATLAB 6.1 and SPSS 14.0.

RESULTS

Behavioural Data

During the scanning sessions, subjects performed a task in which they were instructed to determine, in a 2-AFC, if the figures in the image were 'symmetric' or 'asymmetric' in the vertical axis. This task was selected to ensure subjects were attending to the images, evaluating them as a whole and grouping the lines that were oriented differently from the background. Fig 2.3.3 shows the performance of subjects in this task. Mean accuracy was 79.49% ± 1.39% for MF, 77.04% ± 1.55% for ML and 78.52% ± 1.93% for NC, with no significant difference between conditions [$F_{(2,26)}$=1.85, p=0.176], reflecting the fact that subjects were as accurate in their responses in all conditions.

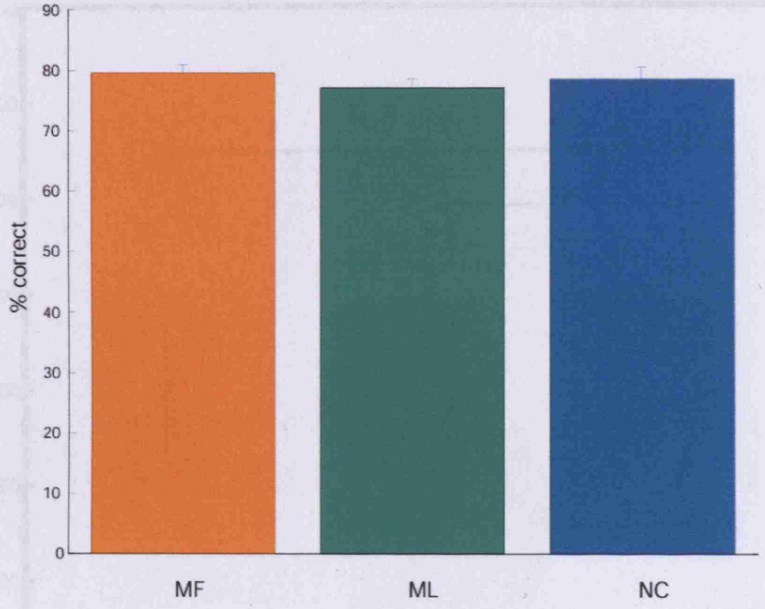


Fig. 2.3.3. Performance in the experimental task. The figure shows the percent of correct responses for each experimental condition averaged across subjects and sessions (mean ± s.e.m). MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

Average RTs for the task were 1105 ± 69ms for MF, 1083 ± 61ms for ML and 1168 ± 72ms for NC (Fig 2.3.4). A one-way repeated measures ANOVA revealed a significant effect of condition (MF, ML, NC) on the RT [$F_{(2,26)}$=15.83,p<0.001]. Individual contrasts indicated that RTs are significantly higher for condition NC compared to MF and ML ([MF - NC]: $F_{(1,13)}$=19.83, P<0.001; [ML - NC]: $F_{(1,13)}$=18.47, P<0.001), but there was no significant difference between MF and ML ([MF - ML]: $F=_{(1,13)}$3.24, P=0.095). This suggests that figures in the NC condition were more difficult to judge and therefore, to exclude the possibility of finding differential brain activations between conditions due to differences in difficulty, we included in the analysis of the imaging data the log-RT as a modulator of no interest.

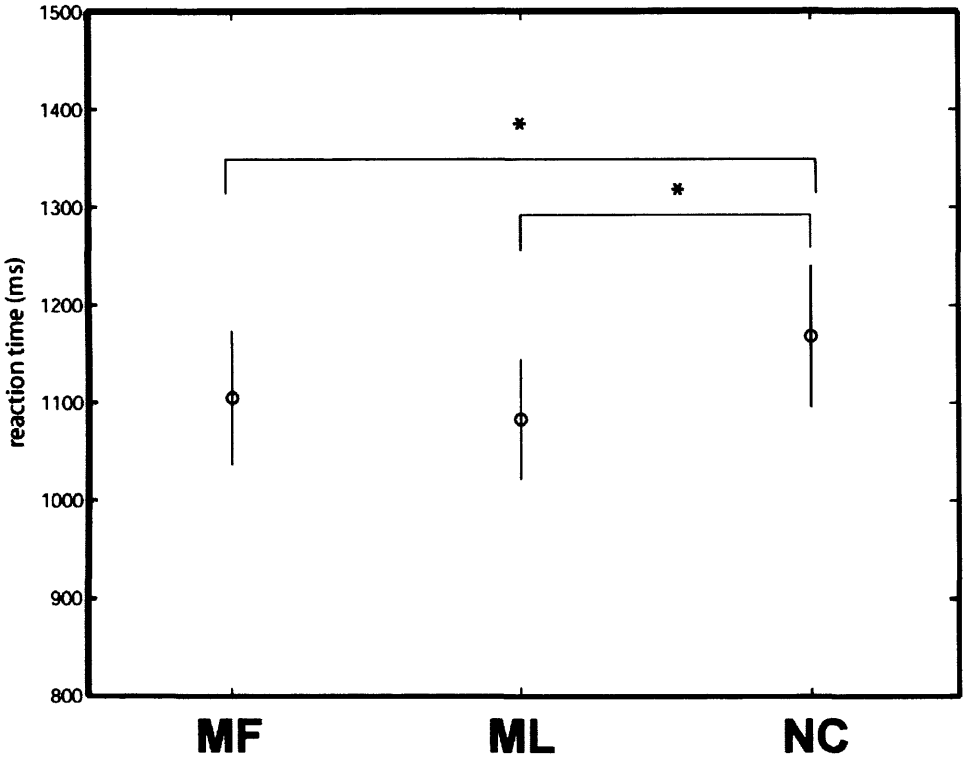


Fig 2.3.4. Reaction times for the experimental task. The figure shows the mean reaction time for each experimental condition averaged across subjects and sessions (± s.e.m). MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

Due to the abstract appearance of the stimuli and differences between individual subjects in interpreting them, stimuli were judged previous to scanning by a different group of subjects, and only those stimuli that were recognized as meaningful more than 70% of the times, and meaningless less than 30%, were included in the MF condition. The opposite criteria were adopted to include stimuli in condition ML and NC. After the scanning sessions, subjects were asked to judge the figures as 'meaningful' or 'meaningless' in a 2-AFC task (see Methods), with the whole set of stimuli being presented over two sessions, with the same presentation and interstimulus time than the one used during the scanning period. To avoid subjects making judgements about meaning during the scanning sessions, they became aware that they were going to perform a different task only at the end

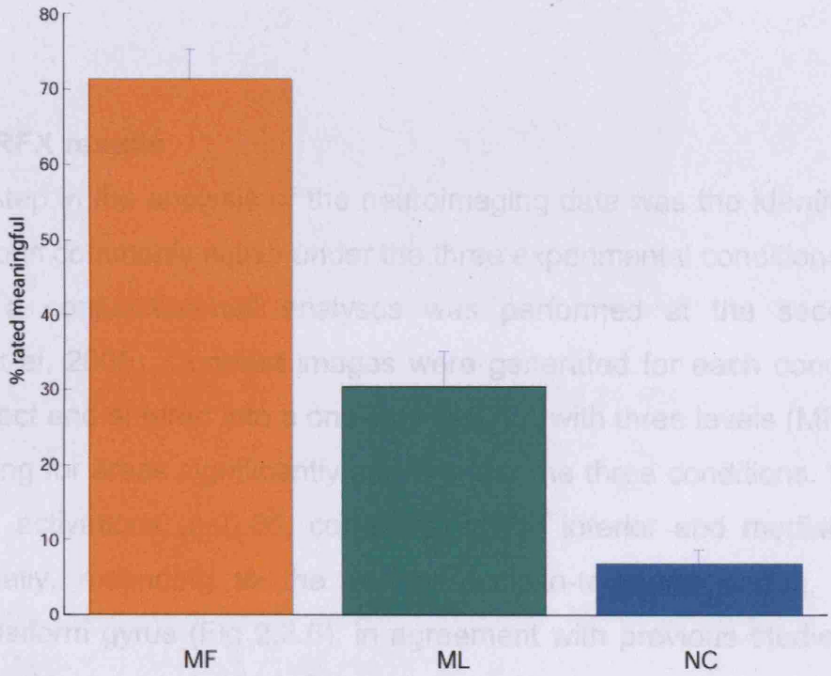


Fig 2.3.5. Differences in semantics between experimental conditions. The figure shows the percent of images in each experimental condition rated as 'meaningful' in a 2-AFC task performed after finishing the experimental scanning sessions. The bars represent the average for each condition across all subjects and sessions (mean ± s.e.m). MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

of the experiment; until that point they believed they would have to perform the same task without been scanned for control purposes. On average, subjects rated as meaningful 71.33 ± 3.97% of the stimuli in condition MF, 30.41 ± 4.66% of those in ML and 6.84 ± 1.86% of the ones in NC (Fig 2.3.5). A repeated measures one-way ANOVA showed a significant effect of condition (MF, ML, NC) on rating [$F_{(2,26)}$=226.06, p<0.001]. Individual contrasts confirmed that, even at very fast presentations like the ones used during the scanning sessions, subjects significantly rate more stimuli in condition MF as meaningful than in condition ML and NC. ([MF - ML]: $F_{(1,13)}$=166.05, p<0.001; [MF - NC]: $F_{(1,13)}$=539.31,p<0.001). The contrast [ML - NC] was also significantly different ($F_{(1,13)}$=53.08, p<0.001), showing that subjects were more prompt to judge as meaningful stimuli that were also collinear. The impact of this result on the interpretation of the neuroimaging activations is discussed below.

Imaging RFX results

The first step in the analysis of the neuroimaging data was the identification of brain regions commonly active under the three experimental conditions. For this purpose, a conjunction-null analyses was performed at the second level (Friston et al, 2005). Contrast images were generated for each condition and each subject and entered into a one-way ANOVA with three levels (MF, ML and NC), looking for areas significantly active under the three conditions. We found significant activations (p<0.05, corrected) in the inferior and medial occipital lobe dorsally, extending to the ventral occipito-temporal cortex, onto the anterior fusiform gyrus (Fig 2.3.6), in agreement with previous studies of form processing (Malach et al, 1995; Price et al, 1996).

We also observed significant activations (p<0.05, corrected), bilaterally, in thalamus, cerebellum, insula and inferior frontal gyrus, and in the left caudate nucleus, as expected in a condition where subjects are engaged in a demanding visual task (Gerlach 1999; Pernet et al, 2004; Lehman 2006).

When evaluating the contrast [MF - ML], we observed significant activations (p<0.05, corrected) in anterior fusiform and posterior inferior temporal/ fusiform gyrus (bilaterally, Fig 2.3.7). Note that activations are somehow

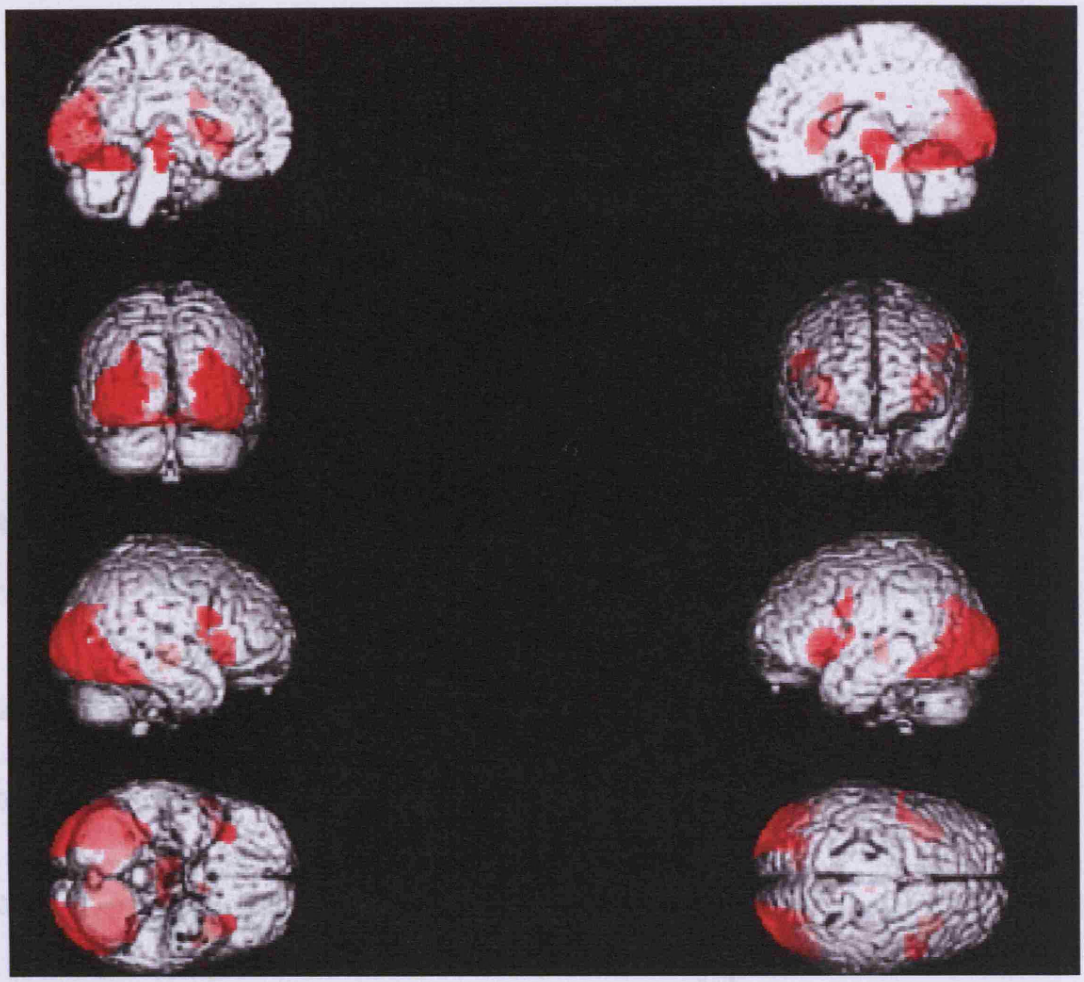


Fig. 2.3.6. Areas commonly activated by all visual stimuli. The figure shows SPMs obtained with a conjunction analysis for [MF, ML, NC]. All activations are shown at a threshold of $p<0.001$, uncorrected.

lateralised, being stronger in the left inferior temporal/posterior fusiform gyrus and in the right anterior fusiform. This contrast not only shows significant activations in visual areas, but also in bilateral inferior frontal gyrus and hippocampus, left inferior parietal and right area 45 ($p<0.05$, corrected). At a lower threshold ($p<0.005$, uncorrected), we also observed right inferior parietal activations, bilateral orbitofrontal cortex and bilateral area 45. These results are in agreement with previous studies that compared real and non-real objects (Price et al, 1996; Vuilleumier et al, 2002). The contrast [ML - NC], that tests for

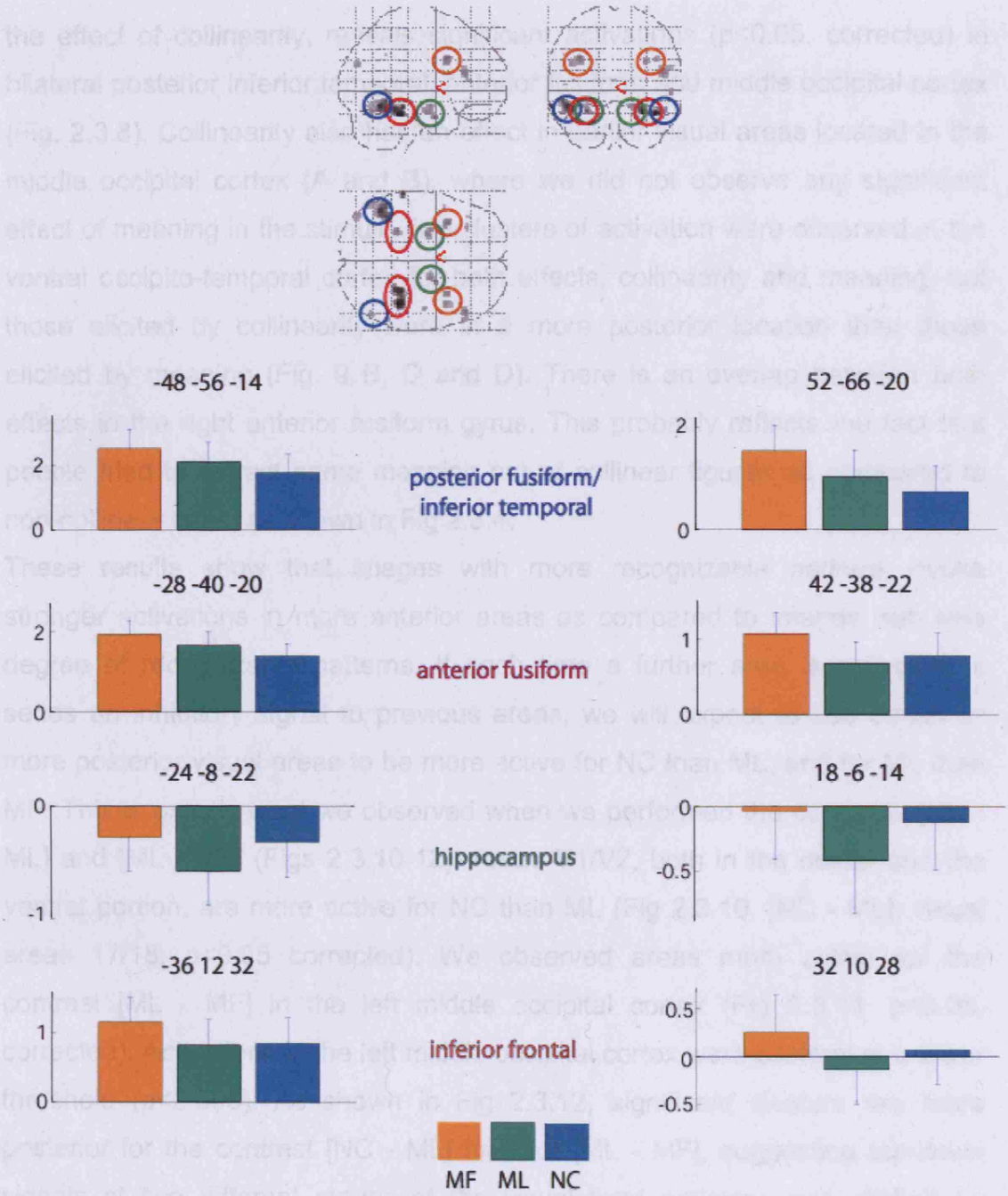


Fig. 2.3.7. Activations evoked by 'meaning' in the visual stimuli. The top panel shows SPMs obtained with the contrast [MF - ML]. All activations are shown at a threshold of $p<0.001$, uncorrected. Group results are shown on SPM's glass brain and the parameters for each condition are plotted from the maxima of the contrast of homologous regions of the right and left hemispheres (percent global signal change ± s.e.m.). MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

the effect of collinearity, reveals significant activations ($p<0.05$, corrected) in bilateral posterior inferior temporal, anterior fusiform and middle occipital cortex (Fig. 2.3.8). Collinearity also has an effect in earlier visual areas located in the middle occipital cortex (A and B), where we did not observe any significant effect of meaning in the stimuli. Two clusters of activation were observed in the ventral occipito-temporal cortex for both effects, collinearity and meaning, but those elicited by collinearity were in a more posterior location than those elicited by meaning (Fig. 9 B, D and D). There is an overlap between both effects in the right anterior fusiform gyrus. This probably reflects the fact that people tried to extract some meaning out of collinear figures as compared to non-collinear ones, as shown in Fig 2.3.4.

These results show that images with more recognizable patterns evoke stronger activations in more anterior areas as compared to images with less degree of recognizable patterns. If each time a further area is activated, it sends an inhibitory signal to previous areas, we will expect to see earlier or more posterior visual areas to be more active for NC than ML, and for ML than MF. This is exactly what we observed when we performed the contrasts [NC - ML] and [ML - MF] (Figs 2.3.10-12). Areas V1/V2, both in the dorsal and the ventral portion, are more active for NC than ML (Fig 2.3.10, [NC - ML], visual areas 17/18, $p<0.05$ corrected). We observed areas more active for the contrast [ML - MF] in the left middle occipital cortex (Fig 2.3.11, $p<0.05$, corrected). Activations in the left middle occipital cortex were bilateral at a lower threshold ($p<0.005$). As shown in Fig 2.3.12, significant clusters are more posterior for the contrast [NC - ML] than for [ML - MF], suggesting top-down signals at two different stages of the visual form pathway, one elicited by collinearity and another elicited by meaning.

It should also be noticed that in the middle occipital cortex there are voxels significantly active for the contrast [ML - MF] and for [ML - NC], but those for the latter are located more laterally (Fig 2.3.13). As well, if we compare the location of the clusters more active for the contrast [NC - ML], it is medial to the significant activations observed with the contrast [NC]. For visual areas that are retinotopically organized, the eccentricity representation in the human brain

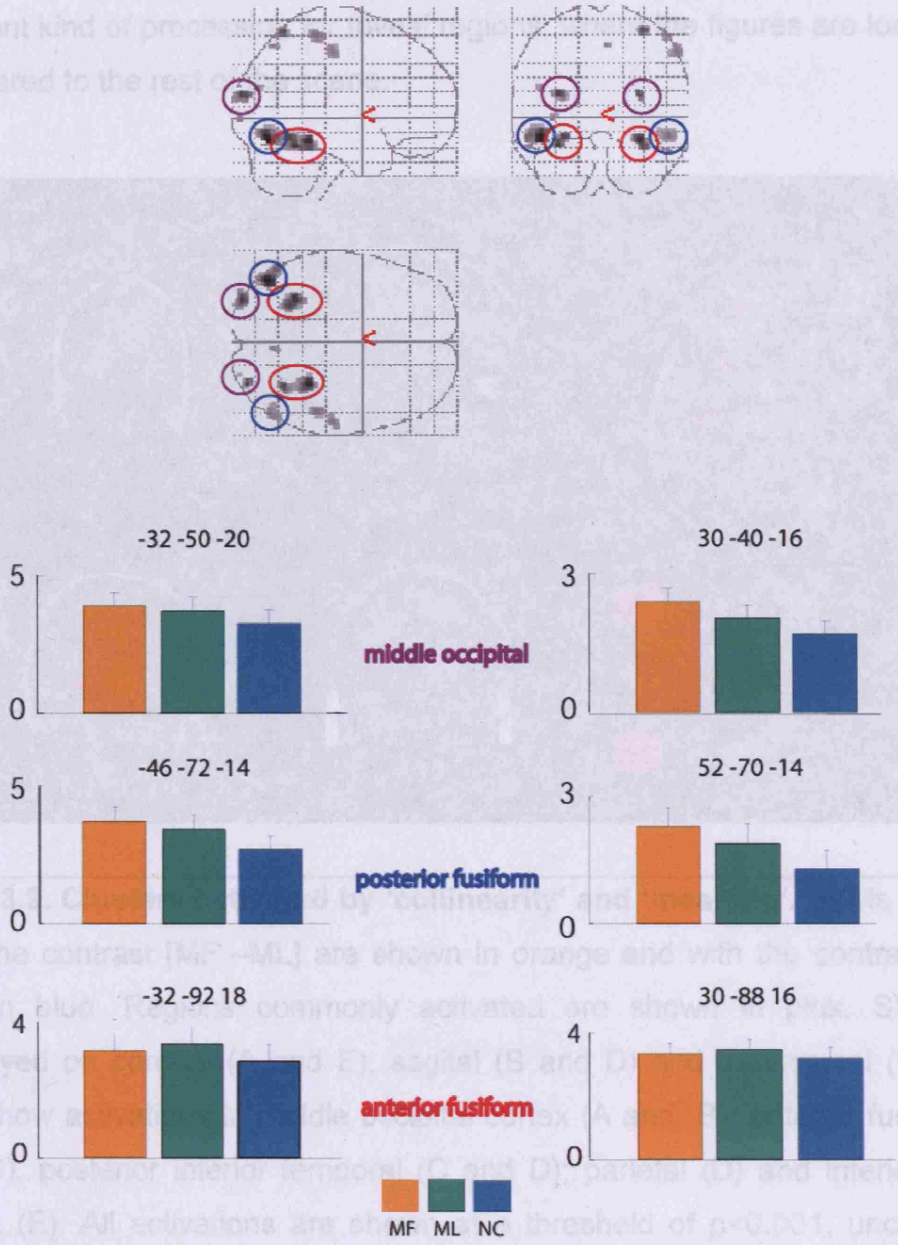


Fig. 2.3.8. Effect of 'collinearity'. The figure shows SPMs obtained with the contrast [ML - NC]. All activations are shown at a threshold of $p<0.001$, uncorrected. Group results are shown on SPM's glass brain and the parameters for each condition are plotted from the maxima of the contrast of homologous regions of the right and left hemispheres (percent global signal change ± s.e.m.). MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

runs lateral-medial (foveal-peripheral) (Wandell et al, 2005), suggesting a different kind of processing for foveal regions, where the figures are located, as compared to the rest of the scene.

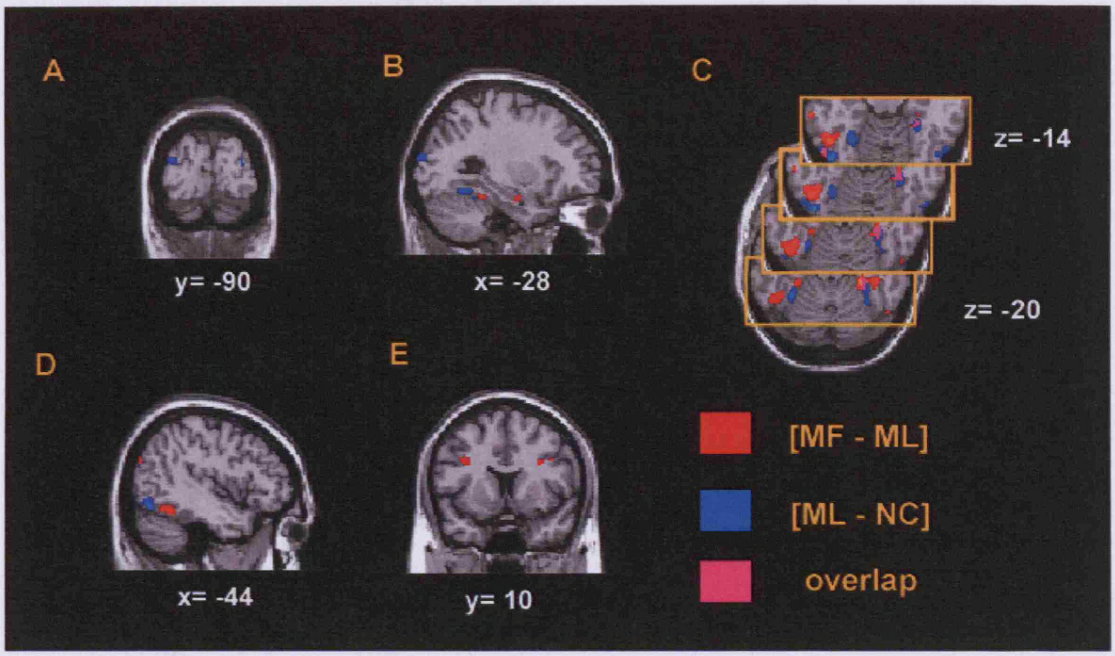


Fig 2.3.9. Clusters activated by 'collinearity' and 'meaning'. SPMs obtained with the contrast [MF –ML] are shown in orange and with the contrast [ML – NC] in blue. Regions commonly activated are shown in pink. SPMs are displayed on coronal (A and E), sagital (B and D) and transversal (C) slices and show activations in middle occipital cortex (A and B); anterior fusiform (B and C), posterior inferior temporal (C and D); parietal (D) and inferior frontal cortex (E). All activations are shown at a threshold of $p<0.001$, uncorrected. MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

Effective connectivity analysis – DCM.

Using a conventional GLM analysis we have shown that images with learnt patterns cause higher activations in more anterior visual areas. On the other hand, when the images contain more random or unknown arrangements, activations are higher in posterior visual areas. This occurs in a step-wise fashion, where condition ML elicits more anterior activations than condition NC,

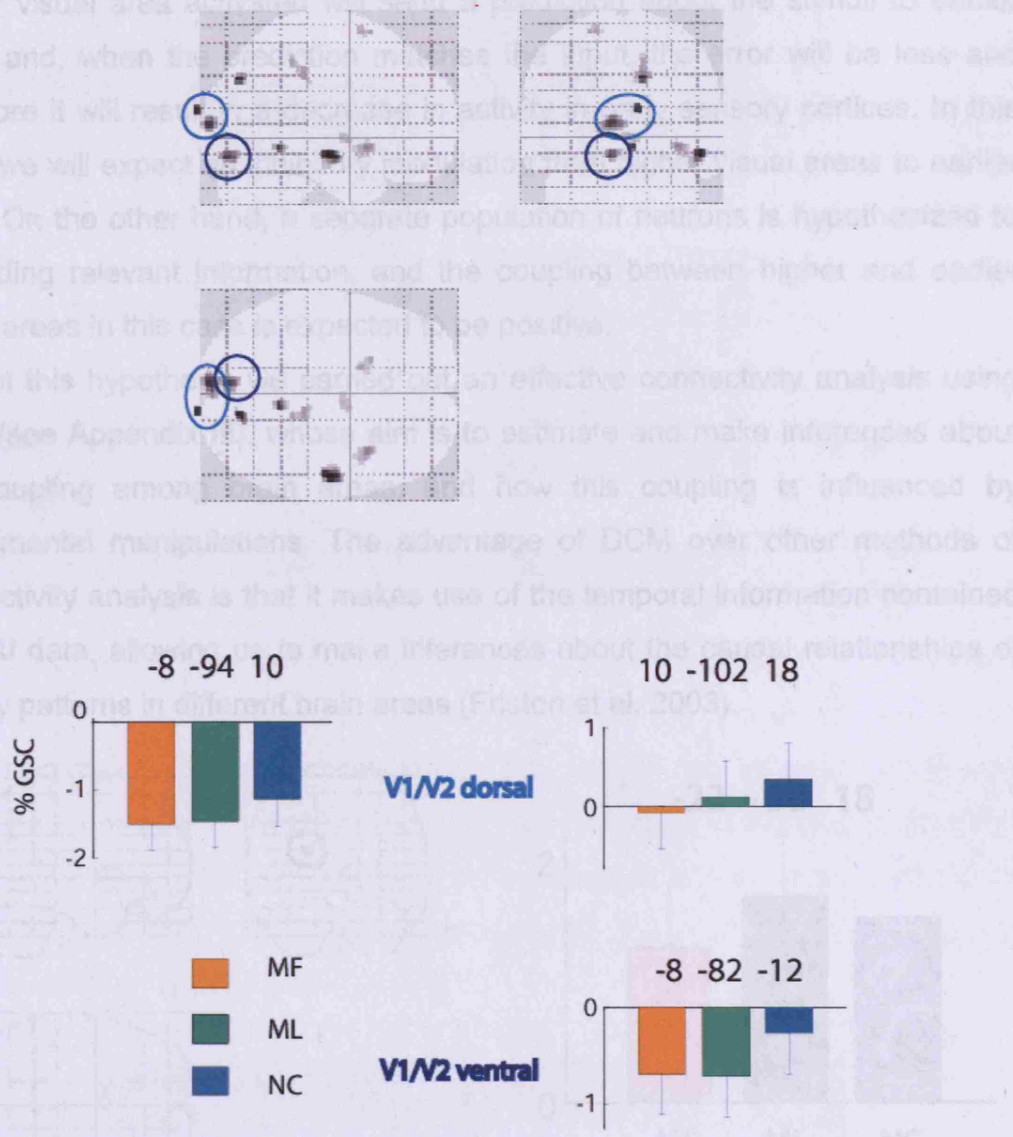


Fig 2.3.10. Brain areas deactivated with the presence of 'collinearity' in the visual stimuli. The figure shows SPMs obtained with the contrast [NC - ML]. All activations are shown at a threshold of $p<0.001$, uncorrected. Group results are shown on SPM's glass brain and the parameters for each condition are plotted from the maxima of the contrast of homologous regions of the right and left hemispheres (percent global signal change ± s.e.m.). MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

and MF even more anterior ones than ML. On the other hand, condition ML activates more posterior regions than MF, and NC activates more posterior ones than ML. These results are in agreement with the hypothesis that each

further visual area activated will send a prediction about the stimuli to earlier areas and, when the prediction matches the input, the error will be less and therefore it will result in a decrease in activity in early sensory cortices. In this case, we will expect an inhibitory modulation from higher visual areas to earlier ones. On the other hand, a separate population of neurons is hypothesized to be coding relevant information, and the coupling between higher and earlier visual areas in this case is expected to be positive.

To test this hypothesis we carried out an effective connectivity analysis using DCM (see Appendix III), whose aim is to estimate and make inferences about the coupling among brain areas, and how this coupling is influenced by experimental manipulations. The advantage of DCM over other methods of connectivity analysis is that it makes use of the temporal information contained in fMRI data, allowing us to make inferences about the causal relationships of activity patterns in different brain areas (Friston et al, 2003).

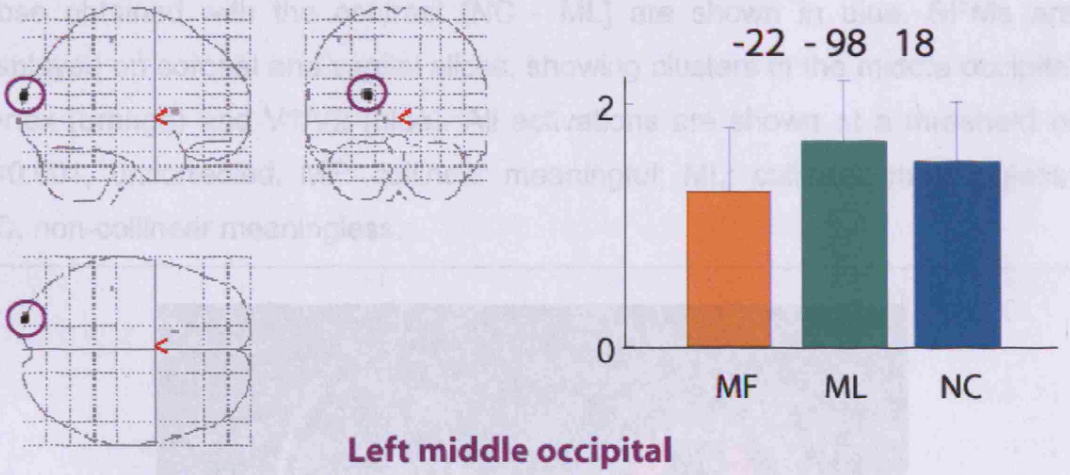


Fig 2.3.11. Brain areas deactivated with the presence of 'meaning' in the visual stimuli. The figure shows SPMs obtained with the contrast [ML - MF]. All activations are shown at a threshold of $p<0.001$, uncorrected. Group results are shown on SPM's glass brain and the parameters for each condition are plotted from the maxima of the contrast of homologous regions of the right and left hemispheres (percent global signal change ± s.e.m.). MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

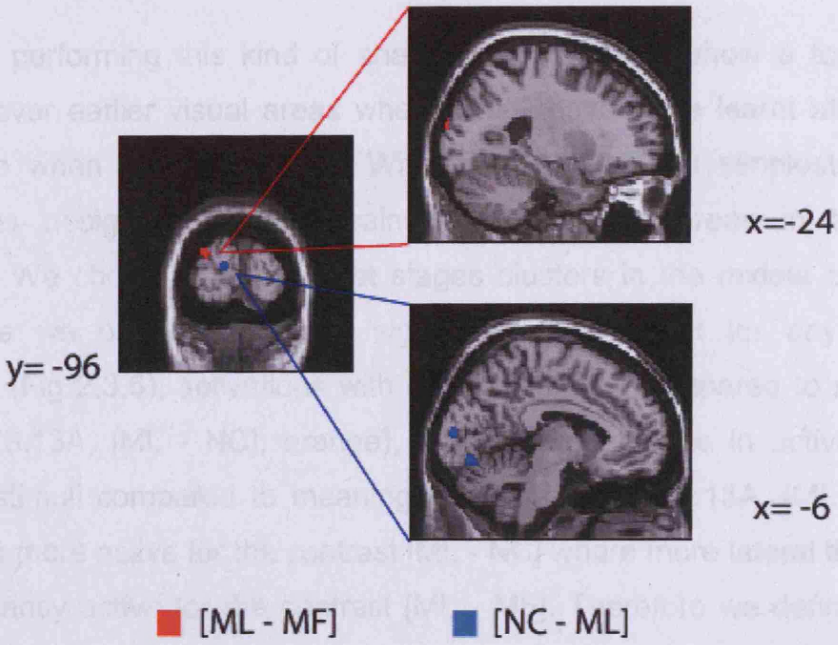


Fig 2.3.12. Deactivations by 'collinearity' and 'meaning' in early visual areas. SPMs obtained with the contrast [ML – MF] are shown in orange and those obtained with the contrast [NC - ML] are shown in blue. SPMs are displayed on coronal and sagital slices, showing clusters in the middle occipital cortex (orange) and V1/V2 (blue). All activations are shown at a threshold of $p<0.001$, uncorrected. MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

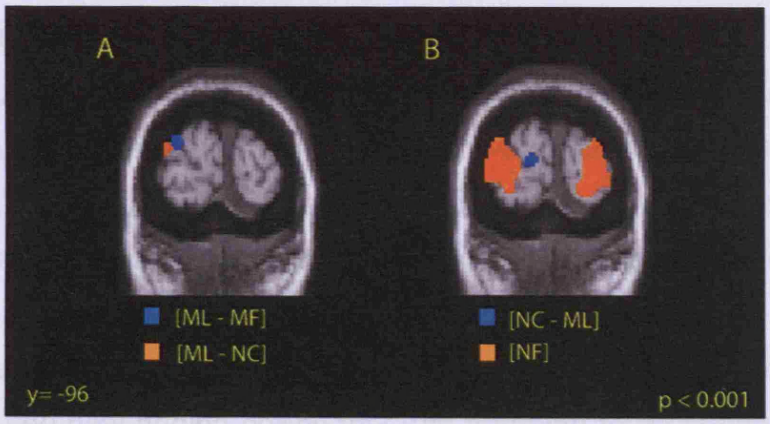


Fig 2.3.13. Lateral and medial clusters in early visual areas. SPMs are displayed on a coronal slice showing clusters in the middle occipital cortex (A) and in V1/V2 (B, blue). All activations are shown at a threshold of $p<0.001$, uncorrected. MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

The goal of performing this kind of analysis here, was to show a top-down modulation over earlier visual areas when stimuli have more learnt attributes compared to when they have less. With this in mind, the simplest model possible was designed. We constrained the DCM to areas in the left hemisphere. We chose as the earliest stages clusters in the middle occipital cortex, since we observed here a significant main effect for any visual presentation (Fig 2.3.6), activations with collinear stimuli compared to random ones (Fig 2.3.13A, [ML - NC], orange), but also a decrease in activity with meaningful stimuli compared to meaningless ones (Fig 2.3.13A, [ML - MF], blue). Voxels more active for the contrast [ML - NC] where more lateral than the ones significantly active for the contrast [ML - MF]. Therefore we defined two areas in the middle occipital cortex: medial and lateral (mMO and lMO), the first one with centre in the maxima for the contrast [ML - MF] and the other one with centre in the maxima for the contrast [ML - NC]. We included as a following stage the region in the inferior temporal cortex (pIT) with the maxima in the whole brain for the contrast [ML-NC] (-46 –72 –14), which did not show a deactivation with MF stimuli. As an area more active for MF than ML stimuli, we chose the more anterior cluster in the inferior temporal/fusiform cortex (aIT), which represented the maxima for this contrast in the left hemisphere. Summarizing, for each subject four regions of interest were defined: mMO, lMO, pIT and aIT (Fig 2.3.14).

Anatomical studies in monkeys have revealed connections between areas in the middle occipital lobe and the ventral occipito-temporal cortex (Seltzer and Pandya, 1978; Distler et al, 1993; Beck and Kaas, 1998; Felleman et al, 1997), and these connections are in general assumed to be reciprocal (Distler et al, 1993; Felleman &Van Essen, 1991). Based on this, we allowed for reciprocal connections between all areas.

We wanted to see how adding collinearity and meaning to an image will result in changes in connectivity between the ROIs. Therefore, we included collinearity (conditions ML and MF) and meaning (condition MF) as modulatory effects, allowing for them to have an effect on any connection of the model. The input to the system was at the level of mMO and lMO.

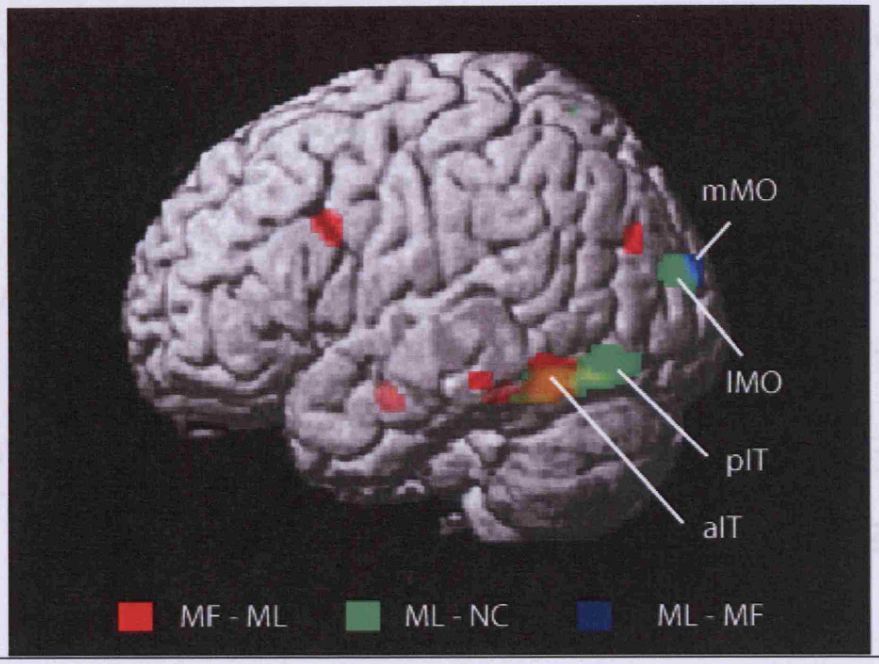


Fig. 2.3.14. Location of areas included in the DCM analysis. mMO: medial middle occipital; lMO: lateral middle occipital; pIT: posterior inferior temporal; aIT: anterior inferior temporal. MF: collinear meaningful; ML: collinear meaningless; NC: non-collinear meaningless.

DCMs were then estimated separately for each subject, and values for each parameter of the fixed connection, modulatory effects and driving inputs were taken to a second-level analysis using a one-sample t-test.

Fig 2.3.15 shows the results obtained with the DCM analysis. Panel A shows the results obtained for the fixed connections. In our model, it represents the connectivity between areas when there is any kind of visual input (i.e. condition NC). There were significant reciprocal connections between lMO and pIT and between pIT and aIT. There was also a significant positive coupling between lMO to aIT, and a negative coupling from aIT to mMO. These results suggest a default circuit between areas mMO, pIT and aIT in the processing of any visual image.

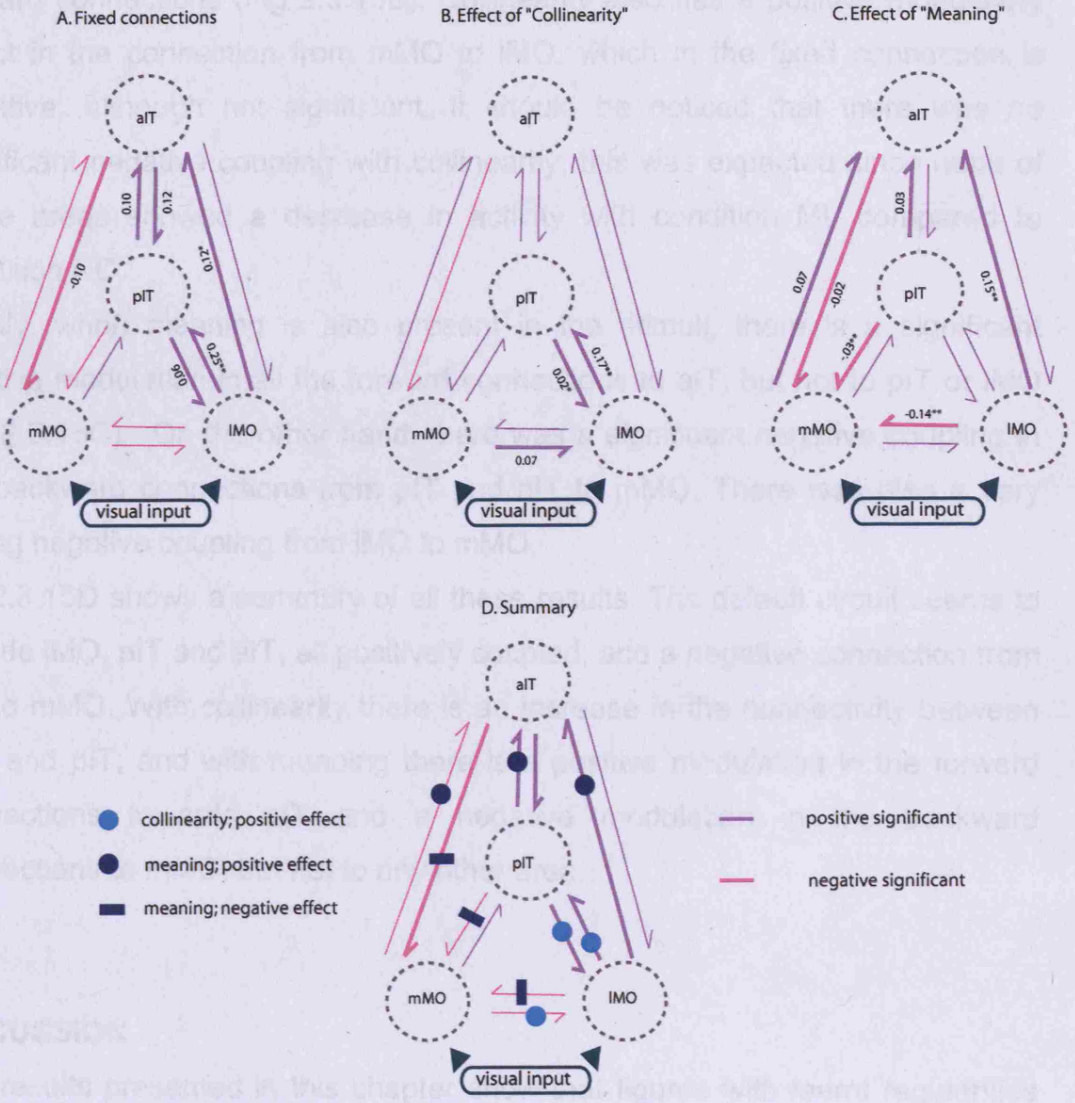


Fig 2.3.15. DCM results. Schematic representations of the group results obtained with a DCM analysis. Each DCM was composed of four areas: medial middle occipital (mMO), lateral middle occipital (lMO), posterior inferior temporal (pIT) and anterior inferior temporal (aIT). Positive parameters are shown in purple and negative ones in pink in all the diagrams. Thick lines represent significant modulations at $p<0.05$, *$p<0.01$ and **$p<0.001$. A summary of all the results is represented in D.

With the presence of collinearity in the visual stimuli, there is an increase in the strength of the connectivity between lMO and pIT, both in the backward and forward connections (Fig 2.3.15B). Collinearity also has a positive modulatory effect in the connection from mMO to lMO, which in the fixed connection is negative, although not significant. It should be noticed that there was no significant negative coupling with collinearity, this was expected since none of these areas showed a decrease in activity with condition ML compared to condition NC.

Finally, when meaning is also present in the stimuli, there is a significant positive modulation in all the forward connections to aIT, but not to pIT or lMO (Fig 2.3.15C). On the other hand, there was a significant negative coupling in the backward connections from pIT and aIT to mMO. There was also a very strong negative coupling from lMO to mMO,

Fig 2.3.15D shows a summary of all these results. The default circuit seems to include lMO, pIT and aIT, all positively coupled, and a negative connection from aIT to mMO. With collinearity there is an increase in the connectivity between lMO and pIT; and with meaning there is a positive modulation in the forward connections towards aIT and a negative modulation in the backward connections to mMO, but not to any other area.

DISCUSSION

The results presented in this chapter show that figures with learnt regularities cause stronger activations in anterior (higher) visual areas and deactivations in posterior (early) ones. With an effective connectivity analysis (DCM) we demonstrated that the addition of learnt regularities generates top-down signals simultaneously and at more than one stage. These can enhance or reduce the activity in an earlier area depending on whether they are targeted at units that are coding the stimulus efficiently or to some which are noisy, overall contributing to more efficiently encode the causes of the percept.

Our analysis shows that all conditions caused activations in visual areas, from the inferior occipital cortex to middle occipital cortex and the anterior fusiform gyrus. This suggests that form information is transmitted continuously and in parallel between visual stages (Zeki and Shipp, 1988; Humphreys et al, 1999),

and not in a discrete fashion, where information will need to be totally processed before being sent to a next stage. Our results are in agreement with previous studies showing a posterior to anterior progression in the processing of object's structural to semantic properties (Vuilleumier et al, 2002; Simons et al, 2003; Gerlach et al, 2002). Stronger responses for meaningful objects were observed in visual areas (bilateral anterior fusiform gyrus and posterior inferior temporal), but also in areas associated with memory, semantics and object decisions, like the hippocampus and inferior frontal gyrus (Gabrieli et al, 1997; Gerlach et al, 1999; Stark and Squire, 2000; Adams and Janata, 2002; Vuilleumier et al, 2002). Areas responding more strongly to collinear stimuli were purely visual, with clusters in the middle fusiform gyrus, posterior inferior temporal and middle occipital. All areas more active for collinear stimuli were always located more posteriorly than those activated by meaningful ones. On the other hand, collinearity and meaning caused deactivations in earlier visual areas, also with a posterior to anterior progression, where the deactivations caused by collinearity are located in V1/V2 and those by meaning were found in the middle occipital cortex, showing that the presence of learnt regularities in an image causes higher activations in more anterior areas which are accompanied by deactivations in earlier ones. These results are in agreement with a predictive coding theory of cortical processing, where higher areas will send top-down signals to earlier ones. The goal of this iterative loop is to reduce the mismatch between this signal and the sensory information, inferring in that way the causes of the sensory stimulation (Rao and Ballard, 1999; Friston, 2003). The activation of further areas specialised in specific functions will contribute with a different kind of signal that could add to those sent by other areas in parallel, each one explaining different attributes of the visual scene.

We tested for changes in top-down and bottom-up modulations using a DCM analysis of effective connectivity. The results obtained showed a positive modulation in the forward connections to pIT with collinearity, and to aIT with meaning. Collinearity also caused a positive modulation in the backward connections from pIT to lMO. In contrast, meaning caused a negative modulation in the backward connections towards lMO. In the human visual brain, the foveal to peripheral organization is represented from lateral to medial region of the brain (Wandell et al, 2005). In our stimuli, the whole arrangement

of oriented lines was centrally presented. Within that arrangement, the figures were always central, and the lines in the most eccentric parts always represented the background. This suggests that positive modulatory signals were sent to populations coding relevant information, in this case the figure, and inhibitory signals to those responsive to irrelevant information, the background.

The DCMs show a strong fixed coupling between lMO, pIT and aIT, in agreement with the fact that we saw activations in all these areas with any kind of visual stimulus, including NC. Again, all these areas seem to be involved in the processing of form, but more anterior ones are more active if the stimulus has some recognizable patterns. It should be noticed that the inhibitory signals come from higher areas, but also from lateral ones, showing that any activated area is inhibiting the other ones. Whereas negative top-down signals came from every higher area, positive ones were only generated from the immediate higher or lateral ones.

To conclude, using a conventional general linear model analysis we have shown that images with learnt patterns cause more intense activations in more anterior visual areas. On the other hand, when the images contain more random or unknown arrangements, activations are higher in posterior visual areas. The results obtained with DCM demonstrated that learnt regularities engage higher visual areas and generate top-down signals simultaneously, at several stages of the system. These top-down signals can enhance or inhibit the activity in the area depending on whether they target units that lead to an overall increase or decrease in synaptic activity. These results are in agreement with the hypothesis that each visual area activated sends a signal explaining an attribute of the stimuli to earlier areas. Top-down signals will contribute to reduce the noise and boost the transmission of relevant information, which will result in a more efficient encoding of the causes of sensory stimulation. If a stimulus has many recognizable patterns the specialised areas processing these patterns will send parallel top-down influences to the area concerned, helping to explain different hierarchical attributes of the visual scene.

Chapter 2.4. – MECHANISMS OF DETECTION AND CLASSIFICATION OF FORMS

INTRODUCTION

In previous chapters I discussed some of the mechanisms of the visual brain which underlie the perception of forms. I have shown that activity in visual areas increases when a scene contains more complex and more recognizable patterns, and that information is refined by recursive loops between early and late visual areas. In these experiments, the presentation times and the quality of the images were adequate enough to allow for perception to always occur successfully. However, there are certain cases in which perception fails altogether – while in other cases, we are aware that a stimulus was presented, but are unable to establish which kind of stimulus it was. In such cases it is said that the stimulus was detected, but it was not classified correctly (although see Grill-Spector and Kanwisher, 2005).

Detection and classification are specific cases of discrimination, which is the ability to tell two stimuli apart (Macmillan and Creelman, 2005). Detection refers to those cases of discrimination where one of the two classes of stimuli contains only the null stimulus (i.e. form vs background). In classification instead, there is not a null stimulus and in general a larger number of stimuli has to be sorted into smaller number of classes (i.e. 100 figures corresponding to two possible classes: lines or polygons).

This chapter studies the neural mechanisms that allow discrimination to occur, in particular the detection and classification of very simple forms.

In any sensory discrimination task, subjects have to make decisions based on some sensory information. These decisions are contingent upon the uncertainty given by the external noise in the sensory input and the internal noise of the system. Discrimination tasks are just a controlled example of the kind of decision-making situations with which we are continuously confronted in the environment.

A common and good framework to study decision-making in the presence of uncertainty is Signal Detection Theory (STD; Thurston, 1927; Tanner and

Sweets, 1954; Greens and Sweets, 1966). SDT acknowledges that decisions depend on the internal noise of the system, the external noise and the response biases of each individual (Heeger, 1997). This becomes more intuitive if explained in the context of a discrimination task. Here I will only discuss the effect of different levels of internal and external noise on the distribution of responses; the response biases are beyond the scope of this thesis. Let us use as an example a detection experiment, where a flash of light is presented in half of the occasions and a grey background in the other half. The task is to respond in a two-alternative forced choice (2-AFC) task "yes" or "no" to the question "did you see a flash of light?" (Fig 2.4.1A). There are four possible experimental outcomes:

- Hit: A flash was presented and subjects saw it (answered "yes").
- Miss: A flash was presented and subjects did not see it (answered "no").
- Correct rejection: There was no stimulus and subjects answered "no".
- False alarm: No stimulus was presented, but subjects answered "yes".

Now, the proportion of trials classified as hit, miss, correct rejection or false alarm will depend largely on the noise of the stimuli (and the internal noise). In a very noisy condition, i.e. where the contrast between the stimulus and the background is very low, there is a high chance of having false alarms and misses, because it is difficult to distinguish between one or the other. We can plot this in terms of a probability function, with the x-axis representing the amount of signal evoked by a flash of light. As can be seen from the graph in the left part of Fig 2.4.1A, there is an overlap between the distribution of the signal evoked by the flash of light and the null stimuli, resulting in a considerable amount of false alarms and misses. On the other hand, if the contrast between the stimulus and the background is high, the distribution of the signal will result in two distinguishable populations (Fig 2.4.1A, right), which will make the judgement easier and therefore result in a lower proportion of false alarms and misses. It would be reasonable to assume that any areas in the brain involved in the perception of these flashes would have a pattern of activity that follows that of the SDT profile. Using stimuli at threshold level and fMRI to measure brain activity, Ress and Heeger (2003) showed that this is the case in early visual areas, with activity corresponding to the report of the subjects. However, if the stimulus contrast is high enough to allow for detection

to always occur, but differences in report arise by changing the demands of iconic memory, noise is injected into the system later on (Hulme, 2006). In this case, the report does not relate to the initial perception, and early visual areas do not follow the report but the physical presence of the stimuli. The same is observed with single-cell recordings in the somatosensory system of macaque, where it has been shown that the strength of the covariations between neuronal activity and perceptual judgments progressively increases from early sensory to frontal areas, demonstrating that the neuronal correlates of the reported experience gradually build up across the system (de Lafuente and Romo 2005; 2006). The state of the system previous to the arrival of the stimulus will also have an effect on the report. Supèr et al (2003) have shown with single cell recordings in macaque V1, that the amount and correlation of activity previous to the presentation of the stimuli determines if a target will be detected or not.

The distribution of responses according to SDT can also be obtained with a classification task. In this case, the task consists of discriminating between two categories of objects, for example houses and faces. We can plot the response probability function in terms of "face signal". In this task subjects will be answering with a 2-AFC the question "did you see a face?" (Fig 2.4.1B). The outcomes here will be the following:

- Hit: A face was presented and subjects answered "yes".
- Miss: A face was presented and subjects answered "no".
- Correct rejection: A house was presented and subjects answered "no".
- False alarm: A house was presented and subjects answered "yes".

Areas involved in face perception, like FFA, follow this signal distribution, whereas those involved in the perception of houses, like PPA, will follow the opposite one (higher responses for "correct rejections", although we can obviously set a different distribution with the question being "did you see a house?") (Grill-Spector et al, 2004; McKeef and Tong, 2007). Here, since the task involves more complex objects, differences are observed in later areas specialised in the processing of that specific category. Therefore, it seems that activity increases in those areas which are involved in detecting that property of the image that make the stimuli different from each other. In the case in which the difference between the stimuli is their physical presence in the scene,

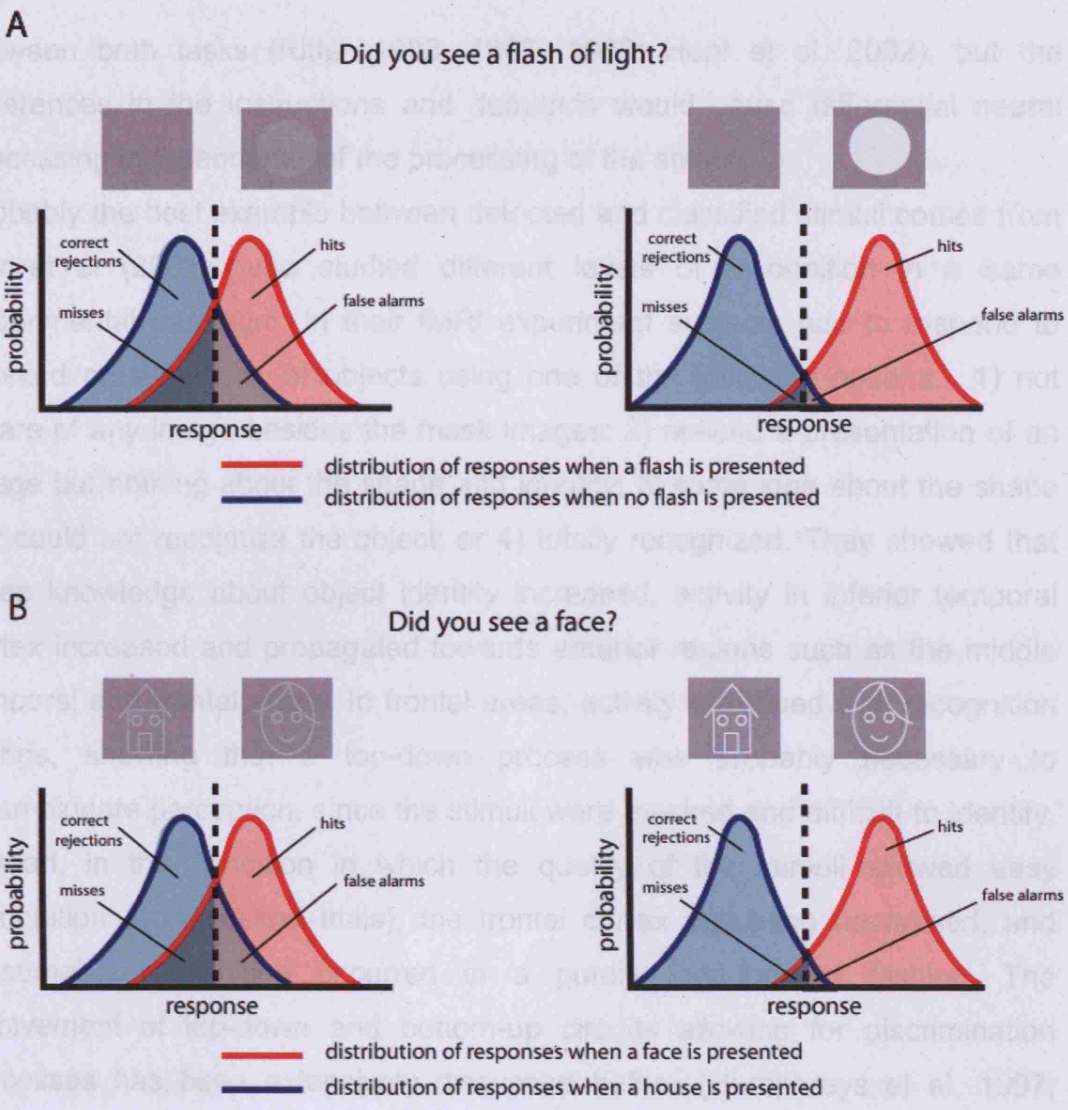


Fig. 2.4.1. Signal Detection Theory (SDT). Distribution of responses according to SDT in a Detection (**A**) and a Classification (**B**) task with high (left) and low (right) levels of noise in the stimuli.

differential responses are observed in early visual areas. Instead, in the case of a classification task, in which the stimuli is always present, responses in early visual areas are the same for both stimuli (assuming there are not differences in basic features between the two stimuli). However, in higher visual areas, like those which are specialised in recognizing a specific category, responses follow a SDT-like distribution.

All these studies are a good example of how it is possible to relate the level of activity of a particular brain area to the decisions of subjects, but they all tackle either detection or classification. Other studies have compared activations

between both tasks (Ritter 1982, 1983, 1988; Hopf et al. 2002), but the differences in the instructions and demands would cause differential neural processing independently of the processing of the stimuli.

Probably the best example between detected and classified stimuli comes from Bar et al (2001), who studied different levels of recognition in a same experimental paradigm. In their fMRI experiment subjects had to respond to masked presentations of objects using one of the following options: 1) not aware of any image besides the mask images; 2) noticed a presentation of an image but nothing about the shape and identity; 3) some idea about the shape but could not recognize the object; or 4) totally recognized. They showed that when knowledge about object identity increased, activity in inferior temporal cortex increased and propagated towards anterior regions such as the middle temporal and frontal areas. In frontal areas, activity increased with recognition ratings, showing that a top-down process was probably necessary to disambiguate perception, since the stimuli were masked and difficult to identify. Instead, in the condition in which the quality of the stimuli allowed easy recognition (non-masked trials), the frontal cortex activation decreased, and presumably recognition occurred in a purely feed-forward fashion. The involvement of top-down and bottom-up circuits allowing for discrimination processes has been extensively discussed before (Humphreys et al, 1997; Pollen, 1999; Lamme and Roelfsema, 2000; Friston, 2002; Gilbert and Sigman, 2007) and shown to be a non-trivial and controversial issue. It is generally thought that some degree of segmentation of low-level features is necessary for visual discrimination. However, low-level features are inherently ambiguous in the environment (e.g. does an edge belong to a shadow, a figure or a texture element?). Therefore, in order to disambiguate them it is necessary to abstract some information about the global arrangement and identity. But, again, this will not be clear until there is some degree of certainty about the low-level features. For this reason, it is thought that segmentation and recognition cannot progress in a simple bottom-up serial fashion, but have to occur concurrently and interactively within recursive loops, where higher-level knowledge will disambiguate lower-order percepts (see Lee et al, 1998; and Grossberg, 1994 for a review).

The problem is that discrimination is performed so rapidly that there does not seem to be enough time for recurrent processes to occur (Thorpe et al, 1996; Riesenhuber and Poggio, 1999; VanRullen and Thorpe, 2001). Thorpe et al (1996) studied categorization in a go – no go experiment, where subjects had to respond if they saw the target category. They found a frontal negativity at 150ms in those trials where subjects reported the target correctly, and concluded that any visual processing of stimuli necessary for categorization was performed and transmitted to frontal areas prior to that time. Since this frontal negativity was observed so early, they postulated that categorization must be accomplished in one forward sweep of activity through the visual system.

Nevertheless, the fact that the processing necessary for categorization can happen before 150ms does not mean that this is always the case. Murray et al (2006) showed that ERPs differences due to accuracy in a task of illusory contour classification are observed at 330-406ms, with sources in the lateral temporo-occipital junction and the superior temporal sulcus. However, if the task involved the discrimination of real contours, which is an easier task, the effects are observed much before (154-192ms).

It seems then that, depending on the amount of information necessary to perform the task, differences arise at different times, probably reflecting the involvement of bottom-up or top-down mechanisms.

This evidence suggests that the demands of the task, the quality of the sensory information and the internal noise, will determine which kind of mechanism will be used.

We wanted to study the differences in brain activity that allowed a stimulus to be detected or classified. We designed an experiment in which subjects had to perform a detection and a classification task after each stimulus presentation. Since the stimulus was the same for both tasks, we could then identify the brain mechanisms behind one process and the other without having other confounds. Previous experiments have shown that the intention of performing a classification task or a detection task causes differences in the ERP from very early stages (Ritter 1982, 1983, 1988; Hopf et al, 2002). Thus, the importance of designing an experiment where the task demands are always the same, independently of whether subjects can classify or only detect the stimulus. We

used stimuli that were as simple and as similar as possible (e.g. a polygon or its component lines; see Fig 2.4.2), controlled for luminance, orientation, size and retinotopic distribution, just with some characteristic features that separate them into two different categories. We wanted to see if using these very similar and simple stimuli, we still see higher areas involved when it is possible to classify the stimuli and only early areas when the stimulus is just detected.

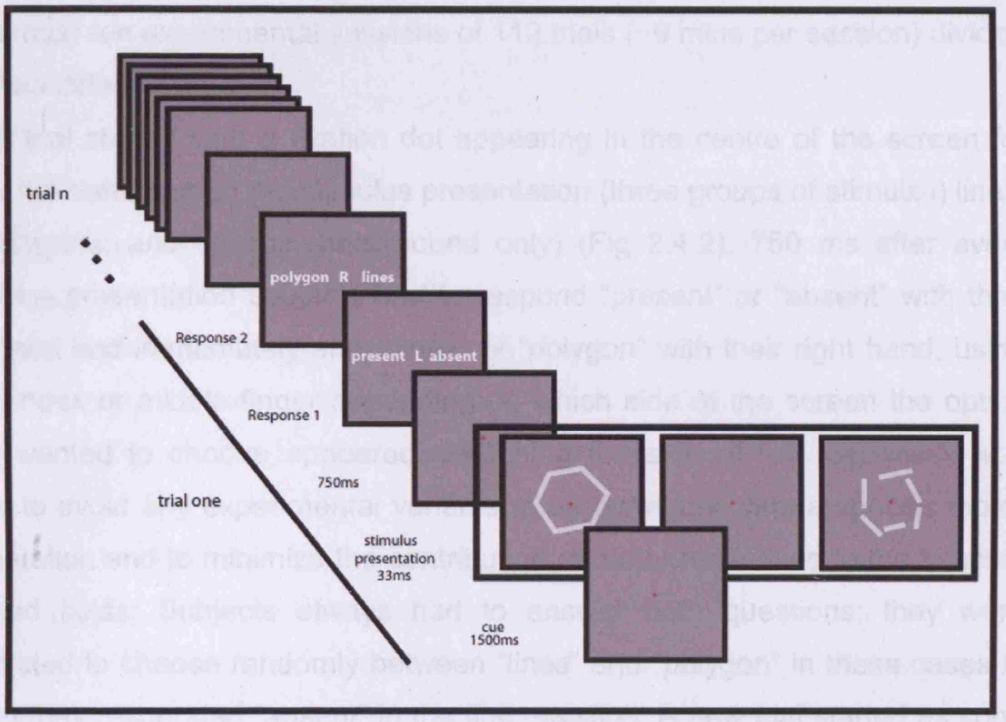


Fig 2.4.2. Experimental design. The figure shows the sequence of events in an experimental trial. After a cue, a stimulus could be presented for 33msec. Subjects had to answer if they saw the stimulus in a 2-AFC task with options "present" and "absent". These options appear in a different side of the screen to avoid any motor preparation. Immediately after this, they had to report if the stimulus was composed of separated lines or if it was a polygon. The options in this case also appear each time in a different side. If they answered "absent" to the first question they were instructed to answer the second one anyway, choosing a random response. "R" and "L" indicate the hand subjects had to use to respond.

We measured brain activity with magnetoencephalography (MEG) since it has an excellent temporal resolution, allowing us to see differences in brain activity from very early time intervals.

METHODS

Stimuli and Experimental Design

Fifteen subjects - 11 male, mean age= 30 (range 20-47, 2 left handed) - performed ten experimental sessions of 112 trials (~9 mins per session) divided into two different days.

Each trial started with a fixation dot appearing in the centre of the screen for 1.5s, followed by a 33 ms stimulus presentation (three groups of stimuli: i) lines; ii) polygons; and iii) null (background only) (Fig 2.4.2). 750 ms after every stimulus presentation subjects had to respond "present" or "absent" with their left hand and immediately after "lines" or "polygon" with their right hand, using their index or middle finger depending on which side of the screen the option they wanted to choose appeared. Switching the side of the responses was done to avoid any experimental variable to be correlated with a specific motor preparation and to minimize the contribution of motor responses to the sensory evoked fields. Subjects always had to answer both questions; they were instructed to choose randomly between "lines" and "polygon" in these cases in which they responded "absent" in the first question. A new trial started as soon as they responded the second question.

The lines and the polygons stimuli consisted of white drawings on a grey background. To avoid subjects attending to a defined region of the stimulus and therefore perform the task based on the detection of a single junction or gap in a specific position of the visual field, we adopted the following strategies: 1) varying the position of the junctions and gaps by rotating the shapes; and 2) including shapes of different sizes, obtaining two groups: bigger and smaller than 3.5 degrees (in total, the stimuli covered between 1-6.5 degrees of the visual field, vertically and horizontally). Both conditions, lines and polygons, were equated for retinotopic distribution, line-length, orientation and luminance (Fig 2.4.3). Only trials with big forms were included in the analysis of the MEG data, since subjects performed at chance level when classifying the small ones.

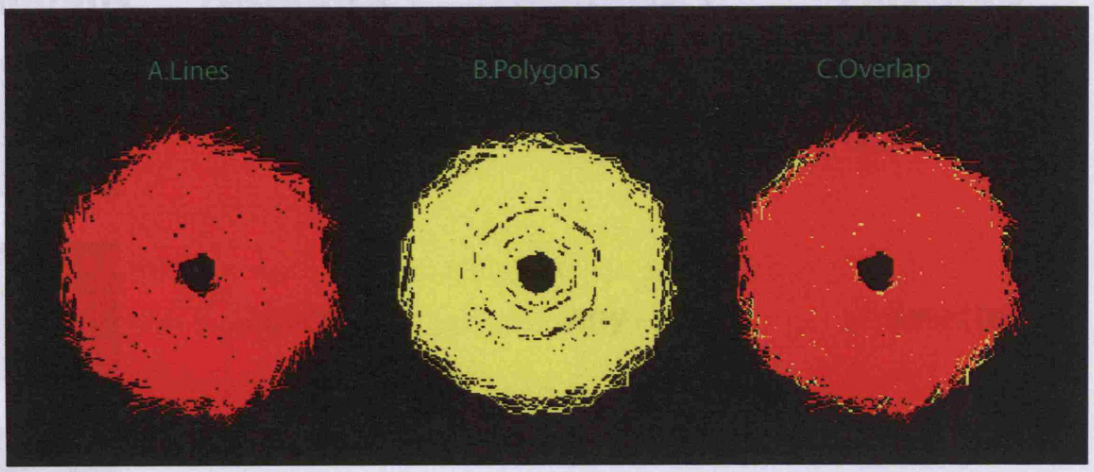


Fig 2.4.3. Stimuli. The figure shows all the stimuli from conditions lines (A) and polygon (B), and the overlap of both (C), to demonstrate that there were no differences in the size, distribution and orientation of the stimuli between conditions. Individual stimulus looked approximately like the examples shown in Fig 2.4.2. Stimuli are shown here in colours for easiness of display; in the experiment they were all white lines and polygons on a grey background.

Trials were analysed according to subjects' response, collapsing across lines and polygons, resulting in the following groups (fig 2.4.4):

- Undetected: trials where a form was presented, but subjects answered "absent".
- Detected: trials where a form was presented, subjects answered "present", but they incorrectly classified them into "lines" or "polygon".
- Classified: trials where a form was presented, subjects answered "present" and they correctly classified them into "lines" or "polygon
- Correct rejections: no form was presented and subjects answered "absent".
- False positives: no form was presented, but subjects answered "present".

Inferences could not be made about the conditions "undetected" and "false positives" since there were not enough trials in these groups. Therefore, results are reported for three conditions: Detected, Classified and Correct rejections. Any difference in the neural activity should be due to the processing of the

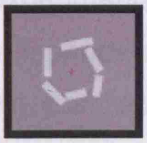
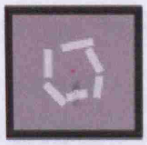
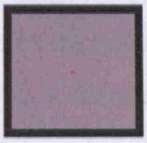
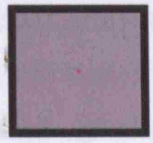
stimulus	response 1	response 2	category
	absent	any	undetected
	absent	any	undetected
	present	polygon	detected
	present	lines	detected
	present	lines	classified
	present	polygon	classified
	absent	any	correct rejection
	present	any	false alarm

Fig. 2.4.4. Post-hoc classification of trials. The figure shows the visual stimulus and the responses subjects had to make for a trial to be included in each analysis category.

visual stimuli, since the task demands in each trial were the same. With this experimental design we could also infer if subjects detected the stimuli, independently of whether they classified them correctly or not. This means that we can assume the incorrectly classified stimuli were detected, because subjects would have answer "present".

At the beginning of each experiment, subjects performed a simple detection task from which their luminance threshold was obtained. The stimuli used for this purpose consisted of a set of configurations of lines and polygons, different from the one in the main experiment. There were between 7-8 repetitions per stimulus type (lines, polygons) and 7-9 different luminance values, depending on subject's calibration. Each trial consisted of a central red fixation-cross presented for 1-2s, after which either one of the stimuli appeared for 33 ms (present trial) or did not appear at all (absent trial). After 500ms, the options "present" and "absent" appeared on the left and right side of the screen, respectively, and subjects had to press a left or right button on a keypad to choose their response. The luminance threshold value (75-80% detection accuracy) was calculated from the psychometric function. From pilot experiments this value was known to be adequate to obtain enough "detected" and "classified" trials.

Before the main experimental sessions on each day, subjects performed a training session where their task was to classify the stimuli into "lines" or "polygon" and received feedback on their performance after every trial. This training session had the same structure as the experimental ones, although it had a different stimuli set and it did not have the "present" or "absent" question.

MEG acquisition

MEG data were recorded at a sampling rate of 480Hz using a 275 third-order axial gradiometers in a Omega275 CTF MEG system (VSMmedTech, Vancouver, BC, Canada) installed in an electromagnetically shielded room. The stimulus was projected with an LCD through a porthole and two mirrors onto a screen situated at 60 cm distance. Responses were recorded through two button-boxes, one for each hand. Trigger events were recorded at every stimulus presentation. Subjects were seated with their head inside the MEG machine and their head position was determined at the beginning and end of

each experimental session by using three coils attached to anatomical landmarks (nasion, right and left preauricular fiducial points). Before the acquisition of the data, a test session was run with a photodiode placed on the centre of the screen to record the exact presentation of the stimuli and the screen delay.

ERF analysis

Analysis of MEG data was performed using SPM5 (http://www.fil.ion.ucl.ac.uk/spm). The data were band-pass filtered at 0.5-45 Hz (Butterworth), divided in epochs starting 100ms prior and finishing 700ms after stimulus presentation, baseline-corrected and down-sampled to 200Hz. Separate averages of trials from different conditions were constructed for each subject after robust average and rejection of trials with eye blinks or other artefacts using amplitude criteria confirmed with visual inspection (1200 ± fTesla (n=10) or 1000 ± fTesla (n=5)). A grand mean for each condition was calculated for all subjects.

Source reconstruction

To make inferences about the functional anatomy of the observed evoked fields, we conducted a source localization scheme for group analysis as implemented in SPM5 (Friston et al, 2007). In brief (Fig 2.4.5), for each subject, a template canonical cortical mesh of approximately 7000 vertices was used, ensuring that activity was reconstructed in the same space over subjects and assuming that subjects have roughly the same head shape as the template head (Mattout et al, 2006). At each vertex, a dipole was placed, oriented normal to the cortical surface. This mesh was coregistered with the MEG sensors via fiducial markers, and a single-shell spherical forward model was created using Brainstorm (Baillet et al, 2001). The forward model for all subjects was then inverted to obtain the amplitude of each dipole within the mesh within the time window of interest in sensor space and source space. The inversion was done under the assumption that the variance at each source can be factorised into source-specific and subject-specific terms. Source-specific covariance parameters are estimated first using the sample covariance matrix in sensor space over subjects and trials using multiple sparse priors. Data from each

subject is then inverted using the empirical priors from the pooled analysis. This procedure ensures that evoked responses are reconstructed in the same subset of sources over subjects. The source estimates were then interpolated from the mesh to the 3D image and, after spatial smoothing, passed to a second-level SPM for classical inference about between-trial effects, over subjects. Only data from the first experimental day was reconstructed, using the fiducial points and sensor locations of the middle session, assuming that the head position was roughly the same through the length of the experimental session.

We also made a source reconstruction of the grand-mean, where we used as fiducial points those for the first session of the first subject.

Fig. 2.4.5. Schematic of the group source reconstruction analysis. Modified from Henson et al, 2007.

RESULTS

Behavioural responses

For each subject, we grouped the stimuli *post-hoc* into the following groups (Fig 2.4.4): undetected, detected, classified, correct rejections and false positives, as explained in the methods section above. To ensure that there was no

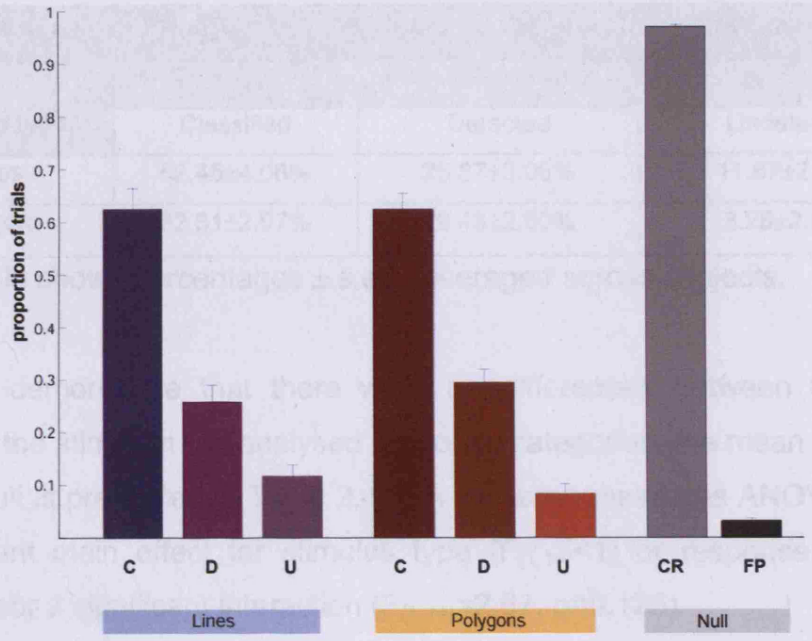


Fig 2.4.6. Proportion of trials in each analysis category. Bars represent mean ± s.e.m. averaged across subjects. C: classified; D: detected; U: undetected; CR: correct rejection; FP: false positive. Results for lines and polygons are plotted separately to show that there were no differences in the pattern of responses obtained for each stimuli group. For the analysis, lines and polygons were collapsed into C, D or U, independently of the stimulus type.

difference in the amount of lines and polygons in each group, we plotted the proportion of these stimuli in each response category (Fig 2.4.6; Table 2.4.1) and conducted a repeated measures ANOVA for Stimulus Type (lines, polygon) and Response Category (undetected, detected, classified). This analysis revealed a significant main effect of Response Category ($F_{(2,28)}$=118.18, $p<0.001$), reflecting that most of the stimuli were correctly classified, but no significant main effect of Stimulus Type ($F_{(1,14)}<1$, $p=0.499$) nor a significant interaction ($F_{(2,28)}<1$, $p=0.628$). This analysis shows that there were no

significant differences in the amount of "lines" or "polygon" stimuli in each response category. In the trials where no stimulus was presented (null stimulus), subjects correctly respond "absent" in 96.89%± 0.51% of the times, with only 3.11% ± 0.51% of the trials being false positives.

Table 2.4.1. Percentage of "lines" or "polygon" in each response category.

Stimulus type	Response		
	Classified	Detected	Undetected
Lines	62.45±4.06%	25.87±3.06%	11.67±2.20%
Polygons	62.31±2.97%	29.43±2.50%	8.26±2.00%

The table shows percentages ± s.e.m averaged across subjects.

To further demonstrate that there were no differences between the basic features of the stimuli in the analysed response categories, the mean perimeter of the stimuli is presented in Table 2.4.2. A repeated measures ANOVA shows no significant main effect for stimulus type ($F_{(1,14)}<1$) or response category ($F_{(1,14)}<1$), nor a significant interaction ($F_{(1,14)}=2.67$, $p=0.125$).

Therefore, differences in the ERF between classified and detected stimuli cannot be explained by the type of stimuli or the amount of pixels in each stimulus, but by the neural mechanisms allowing their correct or incorrect classification.

Table 2.4.2. Mean perimeter of the stimuli in each response category.

Stimulus type	Response	
	Classified	Detected
Lines	370.11 ± 0.94	367.68 ± 1.36
Polygons	367.13 ± 0.69	369.06 ± 2.02

The table shows the mean perimeter of stimuli in number of pixels ± s.e.m.

Event-related fields analysis

We used MEG to study the neuromagnetic responses evoked by the detection and classification of simple form stimuli. Firstly, we averaged each response category for all subjects (grand mean) and then generated two-dimensional interpolated spatial maps of the sensor data at every post-stimulus time point. Inspection of the topographic distribution of the grand-mean for detected and

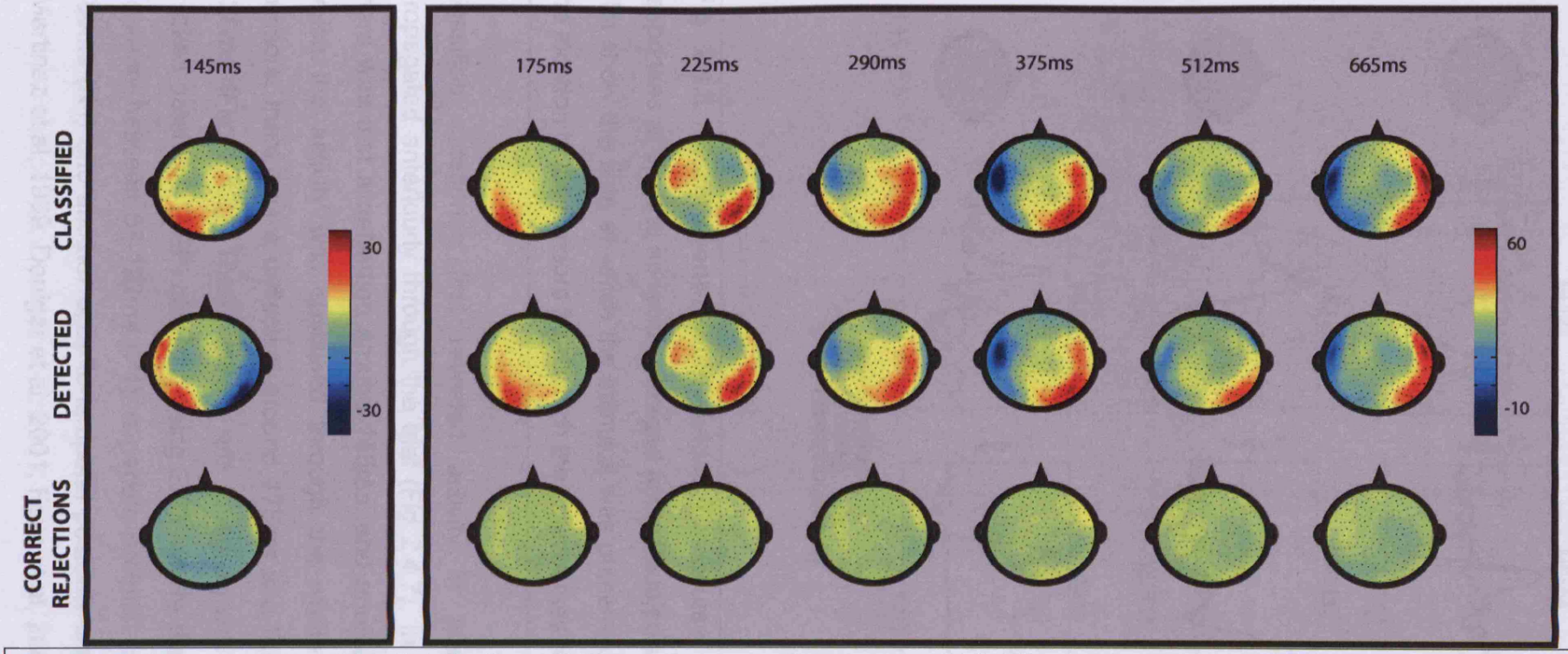


Fig 2.4.7. Mean topographic distribution of the neuromagnetic response. The figure shows interpolated 2D maps of the grand mean of all subjects for each condition at several time points. The colours represent the amplitude of the signal in fTesla.

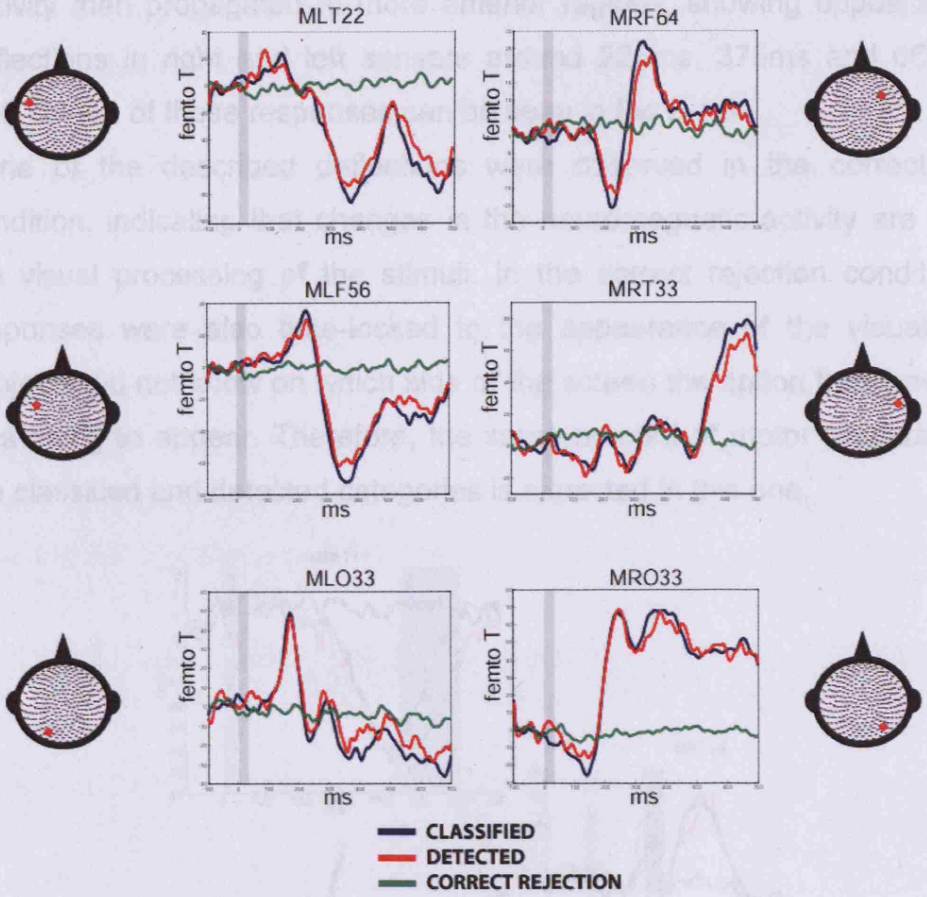


Fig 2.4.8. Mean event-related fields. Timecourse of the neuromagnetic responses at several sensors averaged across subjects. The grey areas in the plot show the time at which the stimulus was presented. The red circles show the location of the sensors from which the timecourses were extracted.

classified categories first revealed activity in posterior regions, which propagated anteriourly through the trial (Fig 2.4.7). In right posterior sensors there was first a deflection around 145ms, and another around 210ms, after which the activity was sustained through the whole trial. In left posterior sensors, there was a deflection around 175ms and then another one around 220ms (Fig. 2.4.8). These results are in agreement with traditional visual evoked potentials (VEP) obtained using EEG, where there is a parieto-occipital positivity between 68-122ms (P1), a parieto-occipital negativity between 124-186ms (N1) and another occipito-temporal positivity between 210-290ms (P2) (Martinez et al, 1999; Doniger et al, 2001; Foxe et al, 2005; Murray et al, 2006).

Activity then propagated to more anterior regions, showing opposite mirrored deflections in right and left sensors around 225ms, 375ms and 665ms. The time course of these responses can be seen in Fig 2.4.8.

None of the described deflections were observed in the correct rejection condition, indicating that changes in the neuromagnetic activity are related to the visual processing of the stimuli. In the correct rejection condition motor responses were also time-locked to the appearance of the visual cue and subjects did not know on which side of the screen the option they have chosen was going to appear. Therefore, the same amount of motor preparation as in the classified and detected categories is expected in this one.

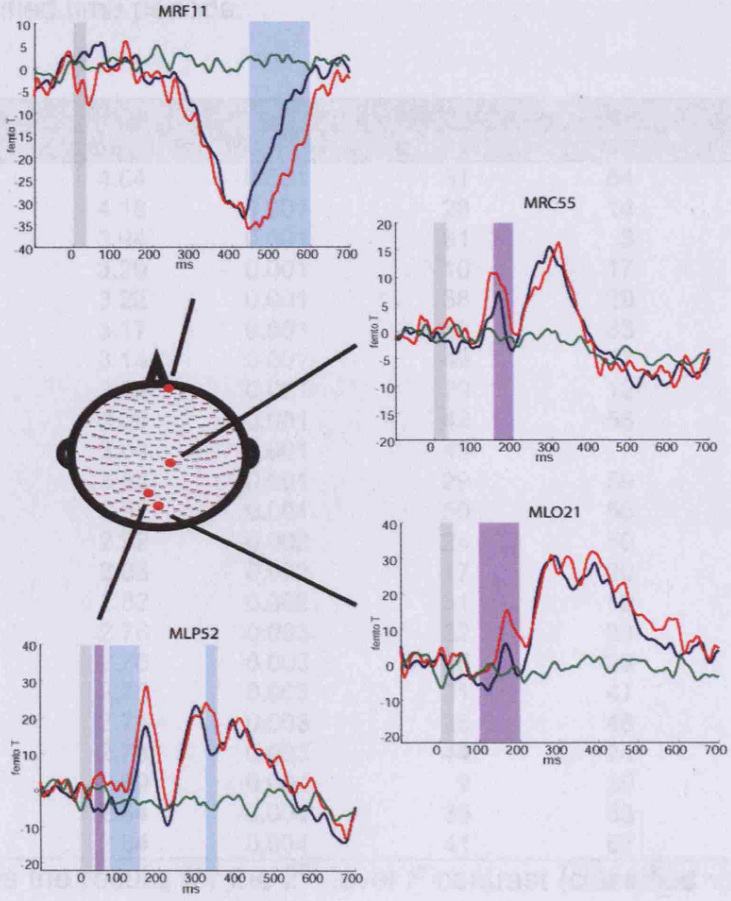


Fig 2.4.9. Differences in the ERF between detected and classified stimuli. The grey areas in the plot show the time at which the stimulus was presented. The blue areas indicate significant differences between conditions "detected" and "classified" at $p<0.05$, corrected. The purple areas indicate differences significant at $p<0.001$. The red circles show the location of the sensor from which the timecourses were extracted.

We then compared the classified and detected categories. For this purpose, two-dimensional interpolated spatial maps of the sensor data at every post-stimulus time point were generated for each subject and then compared between conditions across space and time using a general linear model statistics, testing for experimental differences across subjects (Table 2.4.3). We used the most significant time points of this analysis to define time windows whose length was semi-arbitrarily chosen using visual inspection of the ERF to produce 2-D topographic maps of the average response across the whole time window for each subject. Images were taken to a second-level random effects analysis to localize sensors with significantly different activity across conditions during the specified time periods.

Table 2.4.3. 2nd Level Statistics Results.

F	Z value	P<	x	y	Time
61.36	4.64	0.001	51	54	545
41.84	4.18	0.001	28	14	335
34.50	3.94	0.001	31	3	405
20.25	3.29	0.001	10	17	680
19.08	3.22	0.001	38	29	190
18.31	3.17	0.001	39	33	50
17.82	3.14	0.001	49	37	45
17.13	3.09	0.001	29	12	50
17.11	3.09	0.001	42	56	40
16.50	3.04	0.001	48	7	460
16.49	3.04	0.001	29	59	-50
15.49	2.97	0.001	50	56	45
14.84	2.92	0.002	24	50	50
14.31	2.88	0.002	17	20	-5
13.63	2.82	0.002	31	12	200
12.95	2.76	0.003	32	23	595
12.91	2.76	0.003	25	23	60
12.81	2.75	0.003	31	47	240
12.77	2.74	0.003	25	48	240
12.62	2.73	0.003	44	24	145
12.22	2.69	0.004	9	38	575
11.72	2.64	0.004	35	63	610
11.64	2.64	0.004	41	57	180

The table shows the results for the 2nd level *F* contrast {classified vs detected}. The time column shows the latency of the maxima of the difference between conditions. Columns x and y show the coordinates of a 64x64 2-D topographic maps, where the y axis runs from posterior to anterior and x axis runs from left to right.

Significant differences (P<0.05, FWE corrected) in the ERF between detected and classified stimuli were found between 75-150ms and 325-345ms at

posterior sensors, and between 450-600ms at anterior sensors (Fig 2.4.9). At a lower statistical threshold, but still highly significant ($p<0.001$ uncorrected), differences were also observed between 40-60ms and 100-200ms over posterior sensors, and between 150-200ms in central sensors (Fig 2.4.9). Note that in the case of the 40-60ms time-window the activity of the CR lie in the middle of the tested conditions. This explains, first, why they are significantly different from each other but not from the CR, and also points towards a clear first bifurcation of the stimuli as discussed below.

Differences were not only in amplitude, but also in latency. The latency of the detected condition was significantly longer than those of the classified one at central (MRC55) and posterior (MLP52) sensors between 140-275ms and 245-420ms, respectively (Table 2.4.4). Latencies were not evaluated in the periods including 40-60ms and 75-150ms in left occipital sensors because in most subjects there was not a clear deflection within those time windows in the ERF of the detected condition.

Table 2.4.4. Latencies of ERF deflections within different time windows and sensors locations.

Sensors	**MRC55**	**MLO21**	**MRC55**	**MLP52**	**MRF11**
Time window	**110-185ms**	**95-240ms**	**140-275ms**	**245-420ms**	**350-560ms**
Latency Classified	150.33 ± 5.68ms	163.00 ± 9.43ms	201.66 ± 6.68ms	299.33 ± 9.10ms	445.00 ± 13.76ms
Latency Detected	151.33 ± 7.08ms	166.67 ± 9.46ms	214.00 ± 7.62ms	312.67 ± 11.45ms	456.67 ± 13.93ms
Difference	1.00ms	3.67ms	**12.34ms***	**13.34ms***	11.67ms

The table shows mean latencies averaged across subjects ± s.e.m. *$p<0.05$, obtained with a repeated measures ANOVA. Latencies were calculated as the maxima or minima in activity between the specific time windows for each subject

We could not analyse the data of the undetected condition, nor the lines and polygons separately, because, after artefact correction, there were not enough events in each group category.

Source reconstruction

We first reconstructed the grand mean of the neuromagnetic data for each condition to localize the active sources. The reconstruction of the grand mean shows bilateral sources at 175ms located in the primary visual cortex, superior

and middle occipital cortex and inferior-temporal cortex (Fig 2.4.10, top). At 375ms sources are also at middle occipital regions, but mostly in more anterior regions such as: temporal poles, superior temporal gyrus, postcentral sulcus and cingulated cortex (Fig 2.4.10, bottom).

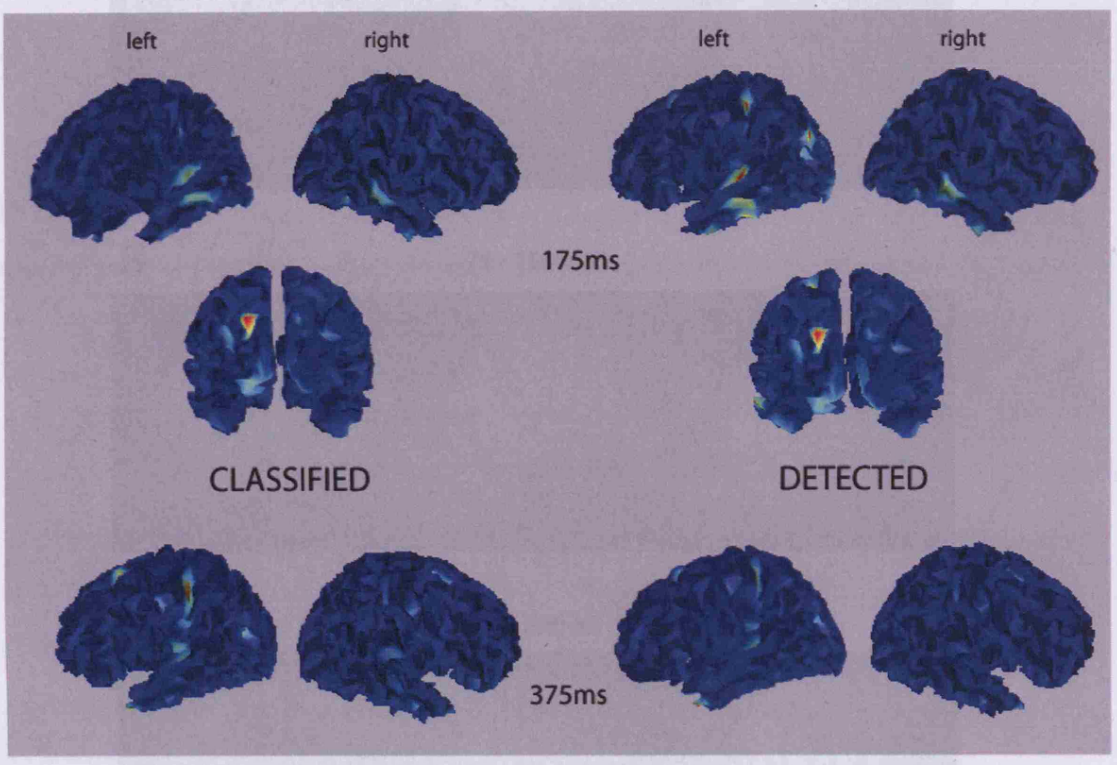


Fig 2.4.10. Mean source reconstruction. The figure shows active sources at 175ms and 375ms for the classified and detected conditions.

To test for statistically significant differences for the comparison of detected and classified conditions between subjects, we reconstructed the anatomic sources for each of the subjects into the standardized space of Talairach and Tournoux (1988) and then tested hypotheses at the between-subject level in source space using SPM5 in the conventional way. Time windows for the source reconstruction were defined from 0 to 600ms at variable intervals based the ERF analysis. Results of the T contrast [detected – classified] show differences in activity of brain sensors in early visual areas between classified and detected conditions at time periods from 0 to 75ms, 75 to 150ms, 150 to 200ms and 200 to 250ms (Fig 2.4.11). Differences were also observed at fusiform areas between 200 and 250ms, although only significant at a cluster level. Significant

differences were not found at any other time window, nor for the reverse contrast [classified – detected].

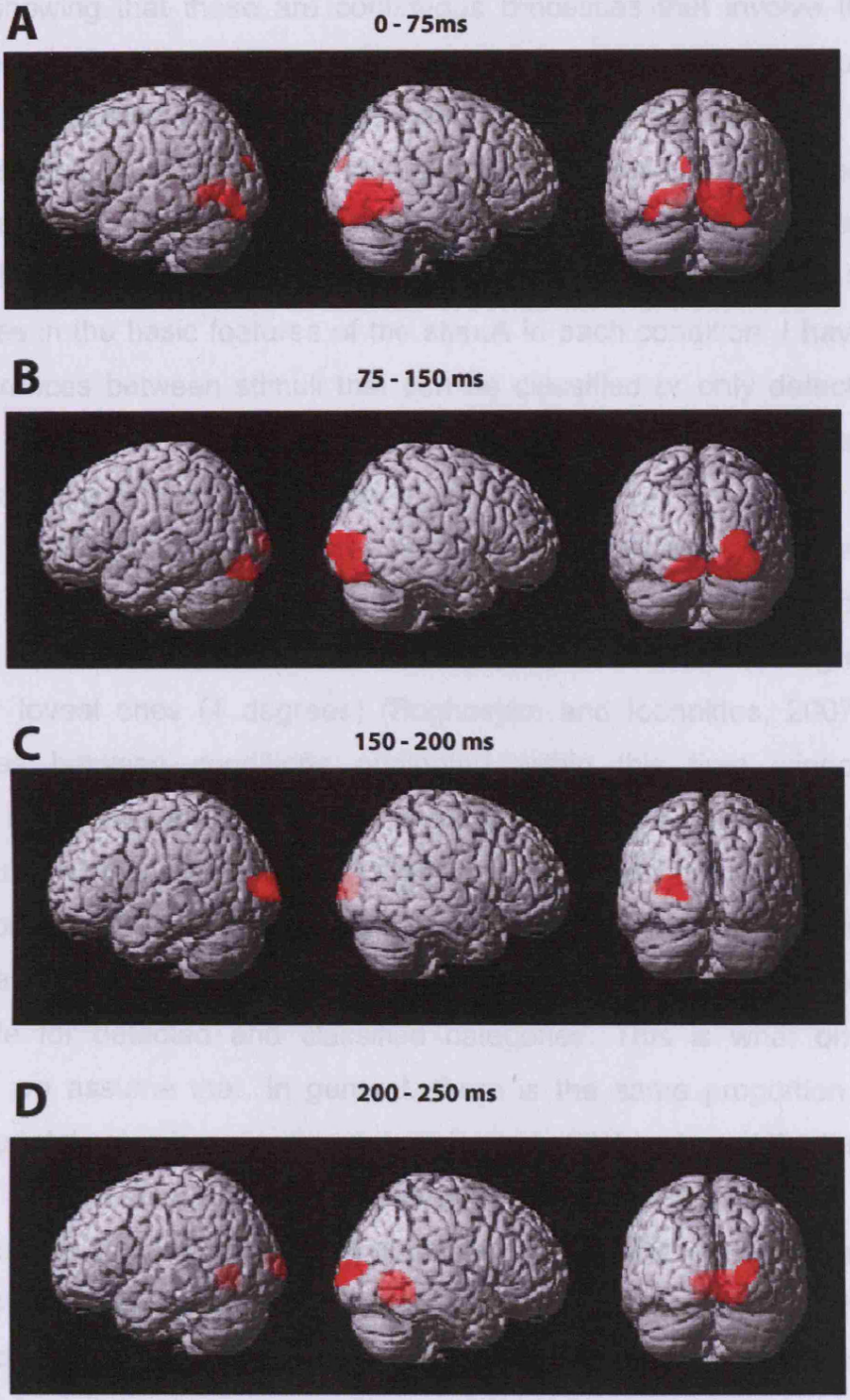


Fig 2.4.11. Group source reconstruction. The figure shows the results of the T contrast [Detected – Classified]. It shows all voxels surviving a significance threshold of $p<0.005$, but only those significant at $p<0.05$ (corrected, SVC) are discussed in the text.

DISCUSSION

In this chapter, I described the differences between classified and detected stimuli, showing that these are continuous processes that involve the same brain areas activated to different extents, and that their successful completion depends on relative levels of signal and noise that are propagated from the earliest stages. With the experimental paradigm used, it was possible to determine the brain mechanisms that allow for classification or detection to occur without confounds due to differences in task demands from trial to trial or differences in the basic features of the stimuli in each condition. I have shown that differences between stimuli that can be classified or only detected arise from the earliest stages in the form pathway, both in terms of the brain areas involved and the processing latencies.

There is a difference in the neuromagnetic activity elicited by classified and detected stimuli from time periods as early as 40 to 60ms. The onset of activity in V1, as measured with MEG, is 41 ms for peripheral stimuli (9 degrees) and 54ms for foveal ones (4 degrees) (Poghosyan and Ioannides, 2007). Since differences between conditions originated within this time window, they probably reflect the amount of noise in the system at the time of the arrival of the stimuli and different amounts of noise transmitted from the retina to the cortex from trial to trial. A result in favour of this view is that the magnitude of the signal in the correct rejections category is intermediate between signal magnitude for detected and classified categories. This is what one would expect if we assume that, in general, there is the same proportion of trials where the total noise in a given area is going to be high or low at the time of the arrival of the sensory stimulation. Those trials in which the level of noise is lower when the sensory information arrives to the cortex would be more likely to be correctly classified than those in which the level of noise is high. Therefore, we expect to see higher noise levels in detected stimuli compared to classified ones. However, in the cases where no stimulus was presented (people correctly reject those trials 97% of the cases in this experiment), we expect equal amount of trials with high levels of noise and half with low levels, with the averaged signal lying in between those for the detected and the classified trials. The internal noise in this experiment, at least in most of the trials, does not impede detection, but does impede classification.

In cases were the signal is sufficiently degraded, like in this experiment, the internal noise will have a big impact on the response of an area, jeopardising its ability to perform the computations necessary for correct classification, and attenuating and disrupting the signal it propagates to other areas. This is seen in the following significantly different time period, between 75-150ms, where the amount of activity elicited by the stimuli was enough to allow for detection, but not classification.

We also observed differences at more anterior sensors at later times (150-200 ms in central sensors, 450-600 ms in anterior sensors), showing that the original uncertainty, reflected in the level of noise, is transmitted through the following processing stages.

The source reconstruction analysis showed differences in source activity in early occipital areas, for early time windows (0 to 75ms) to later ones (75 to 150ms, 150 to 200ms and 200 to 250ms). These early differences showed that the state of V1 at the moment of arrival of the sensory stimulation is critical for the outcome of the processing of the stimulus. This is in agreement with previous findings of neurophysiology in monkeys' V1, where the state of this area previous to the presentation of the stimulus determines if this is going to be detected or not (Súper et al, 2003). In our study, stimuli were mostly detected (and we excluded from the analysis the undetected ones). Therefore the differences we observed in V1 from 0 to 75ms reflect whether the state of activity in V1 at the time of arrival of the sensory stimulation will allow for the stimulus not only to be detected, but also classified. We also observed differences at the source level in V1 at later times, between 75-150ms, 150-200ms and 200-250ms. Due to the late latency of these differences, they probably reflect differences in top-down and lateral processing. Late modulations in V1 cells' responses have also been shown in monkey's single cell responses in the context of a detection trial (Super et al, 2001).

Active sources were also observed at more anterior regions, although without showing significant differences between conditions. Nevertheless, the fact that we see differences in the ERFs at several time points and active sources in posterior and anterior regions for both conditions, suggest that detection and classification represent two different levels of a continuum where information at several levels contribute to task performance, and not two discrete separate

processes. There are not different areas involved in one process or the other – we see the same areas activated to different extents. Particularly, the activity in V1 and the processing taking place in this area seem to be determining the outcome of the task, showing that, at least for this very simple stimulus, there is no further recruitment of areas for the stimulus to be classified correctly. It is possible that if the stimuli involved some more complex categorization, then differences would probably have been observed in more anterior areas.

With these results, we can think of decisions as the outcome of the processing some signal will undergo, the signal being the sum of external sensory signal and internal noise. This processing includes transformation of the signal by lateral, top-down and bottom up mechanisms. If the signal is strong, even with higher levels of noise, the process will be successful. If it is weak, then the level of noise will determine the success of the trial. If the task needs a higher level inference, then the output signal needs to reach a higher threshold. Detection should involve simpler processing than classification and presumably this process can occur even if the level of internal noise is quite high, because to detect a stimulus, it is only necessary to know that there is some external signal, not matter what that is. In the case of classification, the threshold the signal has to reach is probably higher, because it becomes necessary to acquire more knowledge about the stimulus, not only of its presence or absence.

Previous experiments have shown that if the experimental task involve more complex computations, differences between correct and incorrect trials arise later than if they involve simpler ones. In a study of classification (i.e. face vs non-face) and recognition (i.e. Jo vs Anna) of faces, Liu et al (2002) showed that successful classification increased activity at components M100 and M170, whereas successful recognition only increase the M170. Murray et al (2006) showed differences in the ERPs earlier for discrimination of real contours compared to illusory ones. These data suggest that if the task involves easier computations, such as in cases in which it is enough to detect a feature highly diagnostic of the class for general categorization (i.e. faces, buildings) or if objects from class A and B are totally different in every location of the visual field, then differences are found at very early times. However, if it is necessary the integration of information from several parts of the visual field or the stimuli

are quite similar, then differences between correct and incorrect trials are observed later, like it is the case of illusory contours, which are evident only if the whole scene is considered, or processes like face recognition, where it is not enough to detect an eye, but rather it is necessary to integrate the configuration of all features.

In this study, classified stimuli had shorter latencies than detected ones for deflections observed at 201.6ms vs 214ms and 299.3 vs 312.7ms. Previous studies have shown longer P300 latencies for incorrect answers and a component at 150-300ms, related to differences related in object categorization, which covaries with reaction time (Johnson and Olshausen, 2003). These longer latencies are probably reflecting a higher demand to the system, where it ultimately fails to perform the task properly, while in cases where it succeed processing can be stopped earlier. If the signal is strong enough, detection and classification could occur in a forward sweep successfully (Thorpe, 1996). For unsuccessful cases information would probably be attempted to be disambiguated by reducing the noise in a couple of recurrent loops, but if the level of noise is very high, there is not much the system can do to disambiguate the signal.

To conclude, I have provided evidence that suggests that detection and classification are different levels of a continuous process, where the brain mechanisms involved depend on the demands of the task, the external noise in the sensory signal and the internal noise of the system. Situations in which the external noise in the sensory input is too high occur frequently in our everyday environment (object moving quickly, scenes in foggy days). In these particular cases the level of noise in an area at the time the sensory input arrives, is going to determinate the outcome of the classification, and will affect the propagation and further processing of this signal.

Part 3 - GENERAL DISCUSSION

Processing in visual areas and beyond is focused on understanding the sensory information that arrives at the cortex, a strategy used by the brain to acquire knowledge from the world. The brain extracts from the visual scene information about several features: colour, motion, luminance, depth. Each of these features could define a form - a segregated unit or whole. Several forms are grouped together into bigger ones, which are distinguished from the background and associated to concepts and patterns we have learnt before. When a form represents a new pattern, it will trigger a learning process, which after several repetitions will end up consolidating a memory, and result in faster and more educated decisions in subsequent presentations.

The challenge is to understand how the brain encodes and represents all this information. In the overview, I posited a very simple model of form processing with three stages: arrival of sensory input, construction of form and recognition. Subsequent or parallel to these, there are motor outputs, decisions, consolidation of memories and learning. The experiments presented in this thesis study several of the steps of this model, trying to explain how the brain processes different aspects of form perception.

The first experimental chapter deals with the mechanisms of form construction in the brain, i.e. how visual areas assemble all the component parts of a form. The results of this experiment showed that:

- Different steps in the process of form construction activate the same visual areas, the intensity of activity in each area depending upon the complexity of the form.
- Differences between levels of form complexity are more pronounced in intermediate visual areas.
- Visual areas are more responsive if the stimuli fit their RF size.
- The processing of information in the visual areas of one hemisphere is influenced by what it is processed in the contralateral one.
- Most of the voxels in V1 deactivate with form stimuli.

These results suggest that form is constructed in parallel and recurrently within and between visual areas, incorporating excitatory and inhibitory signals, instead of being constructed in a strict hierarchical manner. The intensity of the activity elicited in each visual area will depend upon the complexity of the form located in a particular region of the visual field and will be influenced also by the forms presented in the rest of the scene.

In the second experiment, I studied the brain areas activated by images with different levels of recognizable patterns and the coupling between these areas that will result in top-down and bottom-up mechanisms. The conclusions were:

- More recognizable patterns in an image cause stronger activations in anterior visual areas.
- Images with less recognizable patterns cause stronger activations in earlier visual areas.
- Top-down signals are generated at each stage of the form pathway, where each specialised area can contribute information about different characteristics of the image.
- Top-down signals can be positive or negative, to either enhance or reduce activity in target areas which are coding relevant or irrelevant information, respectively.

We concluded that top-down signals can be generated simultaneously at several stages of the system to either enhance or inhibit the activity in early areas. For a stimulus that has many recognizable patterns, the specialised areas processing each of these patterns will send parallel top-down influences to the area concerned, helping to explain different hierarchical attributes of the visual scene.

Finally, in the third chapter, I dealt with two stages of form recognition - detection and classification - and the differences in brain activity that allows one or the other to occur. The results from that experiment showed that:

- There are differences in the event-related fields (ERF) between detected and classified stimuli since the arrival time of the sensory information to the visual cortex (40-60ms), indicating that the state of the system will have a large influence on the outcome of the discrimination.

- Differences in the ERF between detected and classified stimuli are also observed at longer latencies, which suggests the involvement of top-down and lateral mechanisms.
- The same brain areas are involved in both tasks, showing that detection and classification represent two levels of the same process.
- Differences in source space between both conditions were only found in the early visual cortices, from very early time periods (0-75ms), to later ones (75-150ms, 150-200ms and 200-250ms), suggesting that the processing of simple form in these areas determinates one outcome or the other (i.e. either the form being only detected or also classified).

These results show that detection and classification are different levels of a continuous process, where more iterations and processing of information would be needed in cases in which the sensory information is less, and the internal noise and the demands of the task are higher.

More particular implications of the results of each experiment were discussed in each experimental chapter. Here I will unify them all in a form perception theory for the human brain.

As I mentioned in the Introduction and in Chapter 2.4, the inferences that the visual system will make about a sensory stimulus will depend upon the quality of the signal, the internal noise of the system and the kind of processing necessary to perform the task.

Visual sensory information arrives at V1 and propagates to further visual areas. All conditions in the experiments of Chapter 2.2 and 2.3, independently of the complexity of the form or the amount of extractable patterns, activated early and higher visual areas, showing that the visual information is transmitted through all the stages of the form pathway. Each of these stages processes and tries to extract information from the sensory stimulus.

A pure noise stimulus will evoke a certain amount of unspecific activity in V1 and further visual areas. However strong this activation is, it will be uninformative since there is no pattern that could be extracted from it.

Instead, if a stimulus contains a certain form, for example some randomly arranged lines, each of these lines will be constructed and segregated as an individual entity, although no particular grouping will be performed. Results from Chapter 2.2 show that these simple forms activate all the analysed visual areas

above the level of noise. Now, where the forms are a little more complicated and some are grouped into bigger wholes, they will evoke stronger activity in all the visual areas. This points towards a recursive loop between and within areas, where more neurons will probably have to communicate with each other to put all the parts of more complicated forms together. If form construction were a discrete process occurring only in a forward fashion, where each area was doing a more complex computation than the previous one, we would have seen a pattern of activation where a further area was recruited when there was an increase in complexity. Instead, my results demonstrate that every area, from early to higher ones, activates in response to any visual presentation, with stronger activations observed in more anterior areas if the stimulus contains more extractable or complex patterns.

The construction of a form seems to occur mainly in intermediate visual areas, which is supported by the fact that in the experiment of Chapter 2.2 stronger differences were observed in areas V2, V3 and V3A, than in LO or the VOT; and by the results of Chapter 2.3, where collinear forms activate more strongly areas in the middle occipital cortex than non-collinear ones. Probably, this construction process occurs in the area where RFs fit the size of the stimulus better.

Parallel to this construction process, more anterior visual areas will be performing recognition, as can be observed in the results of the experiment in Chapter 2.3. Recognition in this case refers to the association of a form to a semantic concept, which involves activation of units in higher visual areas responding to all the features of a form, and units in frontal and temporal areas, where those featural patterns would trigger some semantic representations. These featural and semantical stored representations are then going to be compared to the sensory stimulus. Furthermore, a DCM analysis revealed that positive and negative lateral and top-down signals are generated at several stages. A positive recursive loop could boost the transmission of relevant information at each stage, and a negative one can reduce the activity of noisy units. As a result, recursive loops within and between visual areas will refine the processed information.

Construction and recognition of a form are, therefore, not discrete processes, where the second only happens after the first one has finished. My results show

that these are continuous processes, where information is transmitted from one level to the other even before the computations in previous stages are completed. Information will be recurrently processed until the interpretation that best explains the causes of a particular pattern of activity is obtained.

Each time sensory information arrives at the cortex, there will be a first forward sweep that transmits this information from one stage to the other. This will trigger certain representations at each stage, that are not only going to be further processed in higher visual areas, but are also going to be compared with the incoming visual stimulation and with the activity elicited in a particular area, constituting then a recursive loop. After this first forward transmission of information, a sketch of a form will be output from the construction stage, and a sketch of the identity will be output from the recognition stage (Fig 3.1.1). These representations will be sent to further areas and ultimately used to make a decision. In a hypothetical situation in which the system is forced to make a decision ONLY after this sweep, the outcome will be right or wrong depending on the quality of the sensory input and the necessary processing to accomplish a particular task. As shown in the experiment of Chapter 2.4, for stimuli which are very difficult to discriminate, the output of an area will depend very much upon the level of noise that already exists in the system. Therefore, if the level of noise is high, the output from the stage where the extraction of features and the construction of forms occurs is not as clear as in the cases in which the external and internal level of noise is very low. This output will then be processed in further areas and the uncertainty will be propagated through the system. This is the kind of situation in which recurrent loops will help to refine the information – an idea that will be developed below.

The sketches generated at each level of the sensory processing are not only sent forward, but they can also be sent to previous areas for its comparison to the sensory stimulation. This comparison will shape the responses in the lower area, which will then send an updated output, and therefore an updated sketch will be produced in higher areas. These updated sketches are supposed to be better than the original ones generated with the forward information only (Hupe et al, 1998). Therefore, if forced to make a decision based on this information, the system should now be better. If the task involves more complex

computations, then to make a decision it would be necessary more processing, therefore longer latencies of activity will be observed.

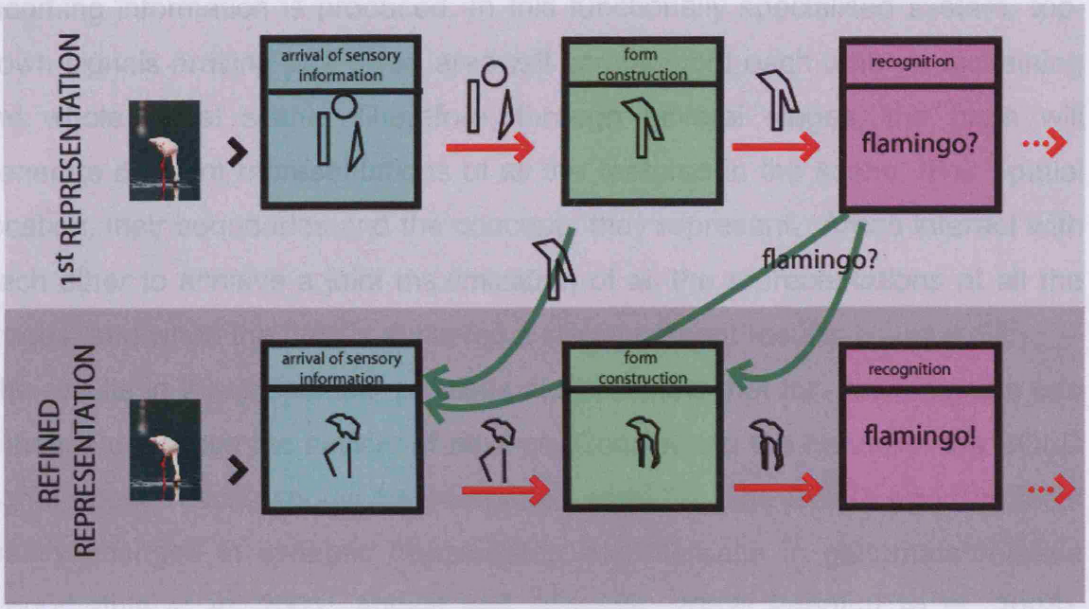


Fig 3.1.1. Model of form processing in the brain. Sensory information arrives at a certain area in the brain in which a representation of the stimulus is generated and passed, via forward connections (red arrows), to the next stage, where another representation will be generated. The representations generated at higher stages can also be sent to early stages via backward connections (green arrows), where these top-down signals will be incorporated to the activity elicited by the bottom-up ones, resulting in a refined interpretation of the stimulus. After some recurrent processing, the representations will match the pattern of activity generated by the sensory stimulation and a stable interpretation will be achieved.

Each area involved in the perception of a form can generate a representation of the specialised function they perform i.e. V2 a small feature, V3A a whole shape, aFus a concept, and comparisons of this information will be performed across levels. What is important to understand is that each area is going to try to generate a representation in response to any visual stimulation; this is

probably the reason why we see activations in all areas with all the stimulus (compared to baseline). Activations are stronger when the specialised areas can actually extract a pattern from the sensory stimulation. This process is repeated in the brain recurrently until a representation that best matches the incoming information is produced. In this functionally specialised system, top-down signals arriving at a given area will complement each other in explaining the whole visual scene. Therefore, through several stages, the brain will generate different representations of all the features in the scene, their spatial location, their boundaries and the concepts they represent. Areas interact with each other to achieve a joint maximization of all the representations at all the stages, and when the best is achieved a stable percept results.

The results in this thesis and previous studies show that top-down signals can enhance or reduce the activity of an area. Considering the nature of the BOLD signal, these results should be interpreted carefully. The BOLD signal reflects mainly changes in synaptic transmission and increase in glutamate release (Logothethis et al, 2001; Attwell and Iadecola, 2002; Peppiatt et al, 2006). Thus, a case in which top-down signals excite interneurons and therefore reduce the output of the area could also result in increased BOLD signal. Independently of the nature of the exact synaptic pattern, they will still represent a change in synaptic activity in that particular area, which in the experiments of this thesis are shown to be mediated by recurrent processes.

In conclusion, I have shown through several experiments in human subjects that form perception is not a discrete process, but rather a continuous one, where information is transmitted serially and in parallel through the levels of a hierarchy of areas performing specialised functions until a stable representation is achieved. At each level, bottom-up and a top-down signals are generated, resulting in recurrent loops that will refine the output of an area and help to encode more efficiently the causes of the sensory stimulation.

Part 4 – APPENDICES

Appendix I. MAPPING HUMAN BRAIN FUNCTION: fMRI AND MEG.

In the experiments presented in this thesis two techniques were used to measure the neural activity underlying the perception of forms: functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG). fMRI has a very good spatial resolution – with this technique it is possible to obtain signals from a couple of millimetres of the brain. Nevertheless, since the measurements obtained depend on changes in cell metabolism and the blood flow (see below), the temporal resolution achieved is very limited. On the other hand, with MEG we measure magnetic fields generated by the neural activity. This is therefore measured as it occurs, resulting in an excellent temporal resolution, but the localization of the signal has to be inferred from the magnetic field recorded from the scalp, which, even in the hypothetical absence of noise, has no unique solution – each magnetic field could haven been generate by an infinite combination of currents originated in different places in the brain.
In this aspect, fMRI and MEG can be seen as complementary techniques, where spatial information is more reliable with the former and temporal with the latter.

BASIS OF fMRI.

fMRI is a technique that allows measures of brain activity in human subjects. It does that by measuring the Blood Oxygen Level Dependent (BOLD) signal, which is proportional to local neural activity, averaged over several millimetres and seconds (linear transform model) (Friston et al, 1994; Boynton et al, 1996; Heeger and Ress, 2002). It was first introduced by Ogawa et al (1990, 1992), and since then it has been extensively used to map cognitive functions in humans.
For fMRI, it is necessary to have a magnetic resonance system that consists of

(modified from McRobbie et al, 2002):

- A magnet that produces a strong magnetic field. A 3T MRI scanner, like the one used in the experiments described in this thesis, has a magnetic field which is 60,000 stronger than the earth's magnetic field.
- Radiofrequency transmitter and receiver coils, which excite and detect the MR signal. The MR signals are produced within the brain in response to radiofrequency (RF) pulses, which are generated by the transmitter coil. They are detected with a receiver coil.
- A computer system for scanner control, image display and archiving.
- Patient couch, comfort and positioning aids.
- Equipment for physiological monitoring and recording of subject's responses (i.e. a pulseoxymeter, eyetracker, keypads, joysticks).
- An intercom to communicate with the subject inside the scanner.

Magnetic Resonance

MRI takes advantage of the magnetization induced in the human body when it is placed in a magnetic field. The strong magnetic field causes H^+ in the human body to align parallel to the main field. The localization of the MR signals in the body to produce images is achieved by generating short-term spatial variations in magnetic field strength across the subject, which are referred as gradients (McRobbie et al, 2002; Logothetis, 2002). The gradient fields are produced by three sets of gradient coils, one for each direction (x, y and z), through which large electrical currents are applied repeatedly in a pulse sequence. The pulse sequence contains radiofrequency pulses and gradient pulses. Each sequence has a TR (repetition time; occurs between the application of one RF excitation and the next one) and a TE (time to echo; time to wait before sampling). During each TR, the excitation and collection of echoes from many slices occur (multi-slice imaging). A radio frequency pulse excites and alters the alignment of H^+. After a the pulse, H^+ get back to their equilibrium position, emitting energy at the same radio frequency at which they were excited. Since a gradient of radiofrequency is applied through space, signals of a particular frequency provide information about the intensity of the signal in a specific point in Voxel space (volume element). Spatial images are therefore obtained by

performing an inverse Fourier transform of the *k-space*, which is the original signal in the frequency domain of a single slice.

The relaxation of H^+ occurs through two processes: a) spin-lattice, which controls the recovery of magnetization along the z axis, with a time constant T1; and b) spin-spin relaxation, that controls the decay of the signal in the transverse axis with a time constant T2. Different tissues give different intensities because they have distinct proton density (number of protons), and different T1 and T2 values. T2 refers to the spin-spin relaxation in a perfectly homogenous magnetic field, which in practice is not the case for any medium, since there are always local magnetic field inhomogeneities. So the relaxation occurs more rapidly, with an effective time constant T2*.

BOLD signal.

BOLD signal is generated due to local inhomogeneities induced by deoxyhaemoglobin (dHb). Hb is composed of two subunits, each one with a polypeptide chain (globin) and haem group (protoporphyrin and iron). The iron atom is the one that binds to oxygen molecules, forming Oxyhaemoglobin. Deoxyhaemoglobin is the form of haemoglobin without the bound oxygen, which leaves the iron in a paramagnetic state (Fe^{2+}). The presence of dHb inside red blood cells generates magnetic field gradients between the blood vessels and the surrounding tissues (Logothetis, 2002). Ogawa et al (1990) first described pulse sequences designed to be highly sensitive to these differences, resulting in signal alterations when the concentration of dHb changes. This can be interpreted as an index of neural activity (see below), and it has been used thoroughly to study brain function in humans.

Neuronal processes measured with BOLD.

Even though the correlation between neural activity and cerebral blood flow has long been established (Mosso, 1881; Roy and Sherrington, 1890; Fulton, 1928; Sokoloff, 1981), the physiological meanings of the haemodynamic changes are still not well understood. In the past seven years, with the widespread use of fMRI as a technique to map brain function, many advances has been made in understanding of this relation. A seminal study by Logothetis et al (2001) , in which the authors performed simultaneous electrophysiological and fMRI

experiments in anaesthetized monkeys, showed that an increase in BOLD response reflects a spatially localized increase in neural activity, which is more correlated with the local field potential (LFP) than to the spikes of single neurons. Since LFP reflects mainly local synaptic transmission mechanisms, they argue that the changes in BOLD signal reflect mainly the input into an area, rather than the output (neurons firing in that area). However, Mukamel et al (2005) showed a high correlation between the BOLD signal and the firing of cortical neurons in humans. Studying patients which were monitored with intracranial depth electrodes for potential surgical treatment, they recorded the activity of neurons in the auditory cortex in response to an auditory stimulus. They also scanned fMRI signals from normal individuals while being exposed to the same auditory stimuli. They analysed these signals with a conventional general linear model (GLM) analysis, using as a regressor the spiking activity recorded from the patients. With this analysis, they showed significant activations in the auditory cortex of the normal subjects in the same region in which they recorded from the patients.

However, the spiking activity of neurons does not necessarily relate to the output of an area. In the cerebellum, for example, where the output is through Purkinje cells only, an increase in the firing rate of these cells do not result in an increase in cerebral blood flow (Thomsen et al, 2004); instead, stimulation of parallel fibres, which results in a decrease in the firing of Purkinje cells, causes an increase in CBF (Mathiesen et al, 1998). Therefore, the interpretation of fMRI studies should take into account the particular anatomy of different regions of the nervous system and the mechanisms through which neural activity causes an increase in blood flow. In a review on the mechanisms leading to increases in CBF, Attwell and Iadecola (2002) suggest that this process has its origins in glutamate-mediated signalling processes locally, and amine- and acetylcholine-mediated global processes. Under this picture, an increase in CBF (or BOLD response) will reflect any kind of neural activity that results in glutamate (mainly) release, this could be any local signal, independently of whether it generates spiking outputs or not. This feat then needs to be taken into account when interpreting fMRI results, but it certainly implies an advantage for the study of processes that change local processing, but not the net output. The main conclusions of all this is that fMRI results

should be interpreted taking into account the local anatomy and physiology of the area under study, since a change in BOLD response can imply different kinds of neuronal activity.

BASIS OF MEG.

Another technique to study the neural processes underlying cognitive functions is magnetoencephalography (MEG), which is based on the fact that any electrical current, such as those generated in the cortex, will produce an orthogonally oriented magnetic field (Fig A1.1). The relationship between electric currents and magnetic fields are formulated in Maxwell's equations, described by the Scottish physicist of the same name. Therefore, using a very sensitive magnetometer, it is possible to measure at the scalp the magnetic fields generated by the electrical activity in the brain.

Spatially structured arrangements of cells are necessary to produce measurable magnetic fields. Excitatory postsynaptic potentials (EPSPs) depolarise the apical dendrites generating differences in charge and a loop of current between the depolarisation point and the soma of the neuron, which are located in different layers of the cortex. The synchronic depolarisation of macrocollums of pyramidal neurons, whose dendrites are parallel to each other and perpendicular to the cortical surface, are believed to be the main generator of the magnetic fields measured in MEG . Magnetic inductions produced by neural currents are of the order of tens of femtoTeslas, necessitating a very sensitive measurement technology. For this reason, MEG uses an array of Superconducting Quantum Interference Devices (SQUID) –an extremely sensitive amplifier that operates at the temperature of liquid helium (4.2K) (Zimmerman and Silver, 1966). The SQUID can be thought of as a very low-noise, high-gain device for transducing magnetic fields or currents into a voltage. It is coupled to a detection coil that senses the magnetic field generated in the brain and a signal coil that transmits the "signals" to the SQUID (Kaufman and Lu, 2003). The output voltage is proportional to the underlying magnetic field and has the advantage that can be amplified and recorded.

Because SQUIDs can operate at acquisition rates much higher than the highest temporal frequency of interest in the signals emitted by the brain (kHz), MEG

achieves good temporal resolution.

The MEG experimental set-up includes:

- A magnetically shielded room, to avoid noise originated from external sources.
- A nueromagnetometer to measure brain activity while subjects perform the experiment. In the experiment presented in this thesis, I used an Omega275 CTF MEG system composed of 275 third-order axial gradiometers.
- A system of coils to determine the position of the subject's head with respect to the sensors. Before the experiment, the coils were attached to the subjects' head in three fiducial points: left and right preauricular and nasion.
- A computer system for control of the equipment, display and storage of the data.
- Equipment for physiological monitoring and recording of subject's responses (i.e. eyetracker, keypads, joysticks).

Once acquired, the data is preprocessed to filter any noise and subsequently epoched into time periods of interest, generally timed around the presentation of the stimuli to analyse event-related fields (ERFs), which are the changes in the magnetic fields generated by the electrical activity of the brain originated in response to the experimental manipulations. The ultimate goal of MEG analysis is to localize the regions where the ERF originated (source reconstruction). To do this, it is necessary to solve the *inverse problem* – to estimate the underlying currents that originated a certain magnetic field. The inverse problem has no unique solution, in principle, each magnetic field has an infinite number of possible interpretations. For this reason, source reconstruction techniques use models of sensor data to make inferences about underlying brain activity (Lutkenhoner, 2003; Kiebel et al, 2008).

Fig A1.1. Magnetic field arising from a current dipole. The recorded MEG signal can be described by the model of a current dipole. The magnetic field arising from a single current dipole has a characteristic topographic pattern – it has two polarity extrema (positive and negative), located in a plane running orthogonal (y) to the axis of the dipole (x). Given a set of currents, it is relatively easy to calculate the associated magnetic field, which is known as solving the forward problem. In MEG, it is necessary to do the opposite – solve the inverse problem, which implies estimating the underlying currents given a certain magnetic field. Modified from Lutkenhoner, 2003.

Appendix II - RETINOTOPIC MAPPING

The knowledge that early visual areas, particularly the primary visual cortex, are retinotopically organized, comes from the study of patients with scotomas due to cerebral lesions during the 19th century. Salomon Henschen (1893) was the principal contributor when he demonstrated that adjacent retinal points are mapped in adjacent cortical points in the striate cortex. It is now well established that in the human brain, the right visual hemifield is represented in the left occipital lobe and the left visual hemifield in the right occipital lobe. In each hemisphere, the upper visual hemifield is represented in the lower part of the brain, ventral to the calcarine sulcus, while the lower visual hemifield is represented in the upper part of the brain, dorsal to the calcarine sulcus.

fMRI allow us to retinotopically map and anatomically define early visual areas (Engel et al, 1994). Using periodic stimuli it is possible to generate waves of neural activity along the length of the occipital cortex. As a stimulus moves from the fovea to the periphery, the locus of responding neurons will vary from posterior to anterior portions of the occipital cortex and each visual field location will alternate between the uniform field and a checkerboard pattern. The timecourse of the alternation depends upon visual field location; peripheral locations are delayed relative to foveal locations. A maximal delay will correspond to the highest eccentricity and minimal delay to foveal regions. The same rationale applies for mapping the visual angle. A wedge stimulus creates a travelling wave of activity moving between the representations of the upper and lower vertical meridian. Because the wedge rotates through the angular component of the visual field, the fMRI signal corresponds to the cortical representation of angular position.

Since the neural activity alternates periodically, the delay can be measured by the phase of the neural activity. With this kind of stimulus, the main parameter of interest for the functional analysis is the delay (phase) of the peak response, not the amount of activity (Engel et al, 1994; Warking et al, 2002).

EXPERIMENTAL PROCEDURE

Stimuli.

We used a slowly expanding ring and a rotating wedge to map retinotopy in the eccentric and angular dimensions, respectively (Fig A2.1).

Stimuli were black and white, maximum contrast, checkerboard patterns, reversing contrast at 4 Hz.

Each wedge or ring session consisted of twelve complete rotations that lasted 30 s each, for a total of 6 minutes per session. Subjects were scanned from 5 to 10 ring and wedge sessions. They performed a fixation task where they had to count how many times a central, small, red dot increased momentarily in size, which occurred between 40-60 times per session at random intervals.

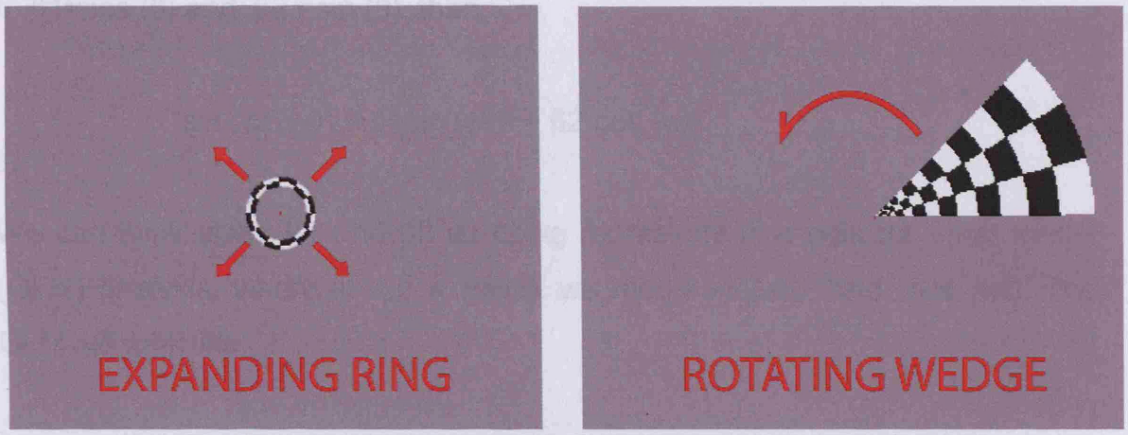


Fig A2.1. Retinotopic mapping stimuli.

fMRI data.

The fMRI analysis procedure was performed in the same way as explained in Chapter 2.1. Each volume was acquired as a series of 28 slices with a TR of 1.82s. The first 8 volumes were discarded to allow for T1 equilibration effects. Images were realigned to the first volume of the first retinotopy session, resliced and coregistered to the anatomical image of each subject. Retinotopic analysis was conducted in a standard fashion (Engel et al, 1994; 1997; Sereno et al, 1994), using SPM2 to calculate phase maps.

Phase maps were calculated using a General linear model ($Y = \beta X + \varepsilon$) in SPM2 (Hutton, 2000).

A particular location in space will be represented by the addition of two angles: one representing the frequency of the stimulus (*w*) multiplied by the time (*t*),

and another angle representing the offset from that circumference (θ, phase of the stimulus).

The sin of the sum of two angles (A and B) follows this expression:

$$\sin (A + B) = \cos (B) \sin (A) + \sin (B) \cos (A)$$

Therefore, for one angle being θ and the other w*t,

$$\sin (wt + \theta) = \cos (\theta) \sin (wt) + \sin (\theta) \cos (wt)$$

If $\beta 1 = \cos (\theta)$ and $\beta 2 = \sin (\theta)$, then

$$\sin (wt + \theta) = \beta 1 \sin (wt) + \beta 2 \cos (wt)$$

We can think about $\beta 1$ and $\beta 2$ as being regressors in a general linear model (GLM) analysis, where in the X matrix we model sin(wt) and cos (wt). The GLM will look like:

$$Y = \beta 1 X1 + \beta 2 X2$$

Y=data

$\beta 1, \beta 2$ = SPM parameter estimates.

X1, X2= Columns of the design matrix representing sin (wt) and cos (wt).

Since,

$$\text{Tan } (\theta) = \sin(\theta)/\cos(\theta)$$

$$\theta = \arctan(\beta 2 / \beta 1)$$

We calculated the inverse tan of the two β images (one representing the parameter estimates for the sine covariate and the other for the cosine), resulting in an image where grey level maps to visual angle. These images were then converted from grey to colour-coded maps for easier visualization.

Structural data and flatmaps.

T1-weighted anatomical images acquired in 176 slices of 1mm thickness, covering the whole brain and cerebellum, were obtained for each subject. Each anatomical image was segmented into grey and white matter using mrVista software (http://white.stanford.edu/software, Teo et al, 1997). Voxels were labelled as one of three tissue types: white matter, grey matter or cerebrospinal fluid. The white-grey boundary of the occipital lobes and neighbouring cortical structures was rendered as a smoothed three-dimensional surface and flattened for further visualisation (Wandell et al, 2000). White matter in this region was manually edited to minimize segmentation errors.

Definition of ROIs.

Colour-coded phase maps of the eccentricity and angular representations data were overlaid onto flatmaps of the cortical surface (Fig A2.2A and B). Foveal representation was identified as the largest clump of contiguous colour or phase. Iso-eccentricity contours (bands of constant phase) run roughly perpendicular to the calcarine sulcus. Eccentricity lines were drawn following the most eccentric band of colour and the external limit of the foveal confluence region.

For each hemisphere, ROIs were drawn by eye on a flatmap of the angular representation with overlaid eccentricity lines (Fig A2.2B) and then superimposed into 3D reconstructions of the cortical surface (Fig A2.2C). V1 was identified as a complete representation of one entire hemifield (180° of phase, half the colour map) centred on the calcarine sulcus (Fig A2.2C). At V1 boundaries (vertical meridian), the orderly progression of colours reverses direction back to the representation of the horizontal meridian, forming a ventral representation of the upper visual quadrant (V2v) and dorsal representation of the lower visual quadrant (V2d). The colour progression returns back to the representation of the horizontal meridian, which determines the limits of areas V3d (dorsal) and V3v (ventral), each one comprising, again, a quarter of the visual field. V3A was identified as a region representing a full hemifield dorsal to V3d. On the ventral side of the brain, a complete hemifield was present beyond area V3v, a region identified as V4 (Wandell et al, 2005). Our maps did not have the definition to allow us to identify other retinotopically organized

regions, like LO-1, LO-2, VO-1 and VO-2 (Brewer et al, 2005; Larsson and Heeger, 2006).

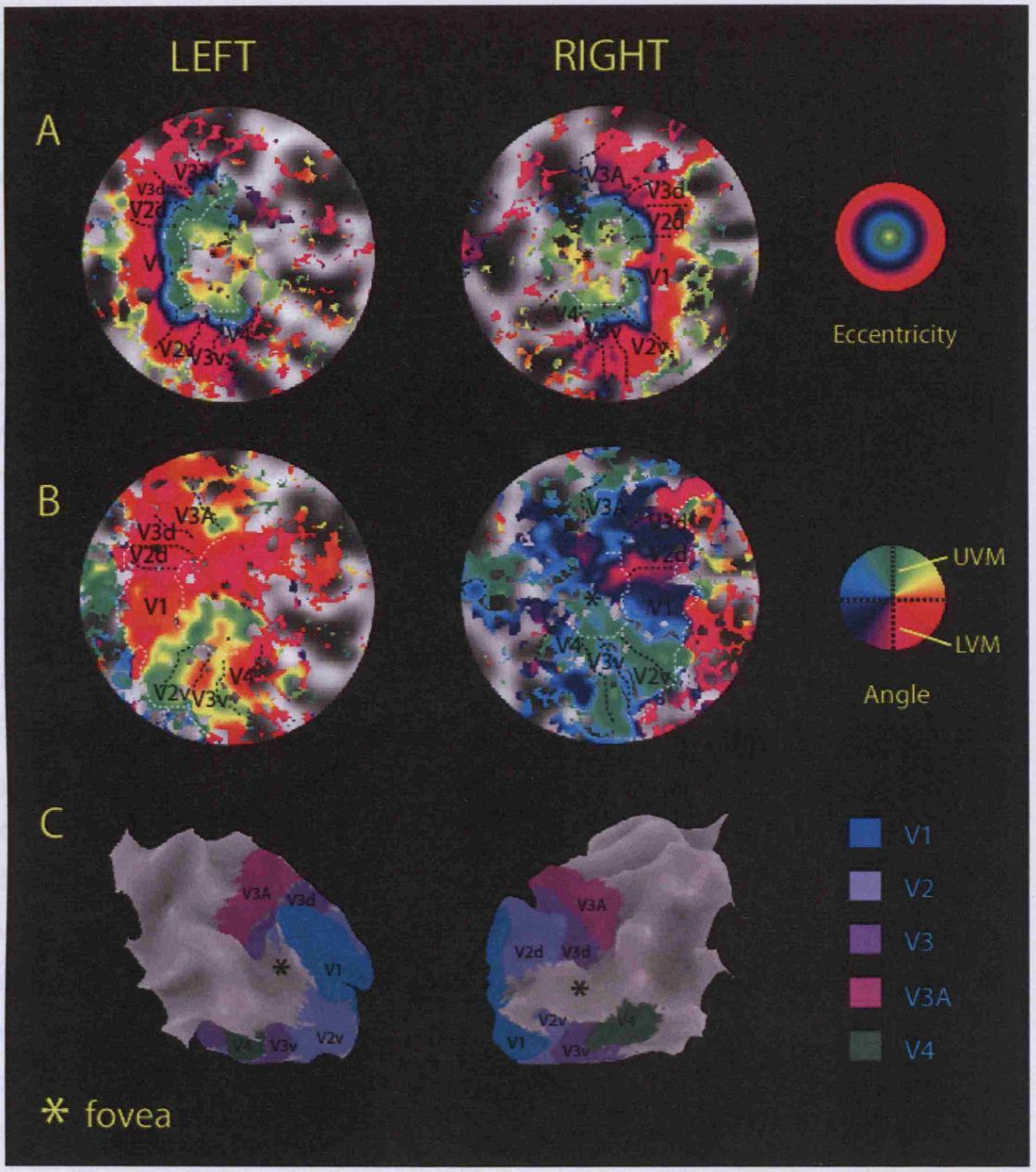


Fig A2.2. Retinotopic mapping. Colour coded visual eccentricity (**A**) and angle (**B**) representation overlaid on flattened cortical maps of the occipital lobe of an individual subject. The white dashed-lines represent isoeccentricity zones. Visual areas were defined by identifying the upper- and lower-vertical and horizontal meridians, which are shown in black dashed-lines on the cortical maps. **C**) Visual regions of interest overlaid into 3D representations of the cortex. The star represents the foveal confluence region.

Appendix III - DYNAMIC CAUSAL MODELLING.

Effective connectivity analyses allow us to study the influence certain components of a system will have over the other components. Models of effective connectivity are appropriate for situations where we have *a priori* knowledge and experimental control over the system of study (Friston, 2005b). Therefore, it is possible to apply this approach to the study of the brain by modelling interactions among neural populations using neuroimaging methods: haemodynamic or magnetic time series. During my PhD, I used Dynamic Causal Modelling (DCM) to make effective connectivity analysis, a technique developed by Friston and co-workers at the Wellcome Neuroimaging Centre (http://www.fil.ion.ucl.ac.uk/SPM).

Main concept

The aim of DCM is to estimate and make inferences about the coupling among brain areas, and how this coupling is influenced by experimental manipulations (Friston et al, 2003). By taking into account the anatomical structure of the system and the interactions within that structure under different experimental conditions, DCM allows us to model brain activity at the neuronal level. This is of relevance because it provides information that is not directly accessible in fMRI. The central idea of DCM is to treat the brain as a dynamic input-state-output system. The inputs are the experimental manipulations and the outputs are the haemodynamic signal measured with fMRI. The state variables comprise the neuronal or synaptic activity and some biophysical variables that determine the output.

In short, with DCM a model of neuronal dynamics is created and then transformed into area-specific BOLD signals using a haemodynamic model of fMRI measurements (Fig A3.1). The parameters of the joint forward model (i.e. the neuronal and the haemodynamic models) are then estimated using a Bayesian estimation scheme to best fit the experimentally observed BOLD response.

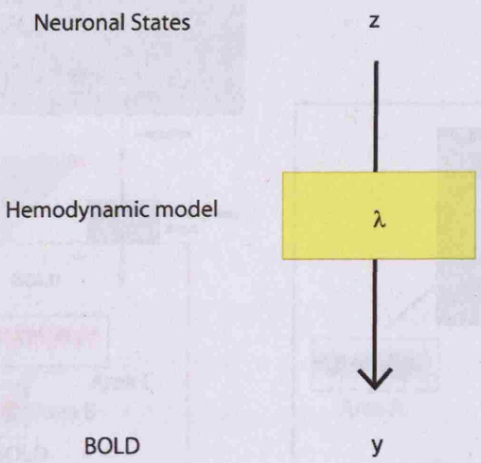


Fig A3.1. Schematic representation of DCM's main concept. The neural dynamics (z) predicted with the bilinear state equation enter a model of the haemodynamic response (λ) to give the predicted BOLD responses (y).

Neuronal dynamics estimation

The first step in DCM is to construct a simple and realistic model of interacting brain regions. This is generally done by extracting the time series of BOLD activity of key brain areas activated during an experiment, which are identified with conventional fMRI analysis (Fig A3.2). Based on anatomical knowledge of the brain, some models of how these areas will be connected and which of the connections or states of the model will be affected by the experimental manipulations, are proposed.

The neural state variables do not correspond directly to any common neurophysiological measurement (such as spiking rates or local field potentials), but represent a summary of neural population dynamics in the respective regions. What is then being modelled is the temporal evolution of the neural state vector

$$\dot{z} = dz/dt = F(z,u,\theta^n) \qquad (1)$$

Fig A3.2. DCM procedure. Modified from Stephan et al, 2007.

Where the state z and the inputs u are time-dependent, whereas the parameters θ^n are time invariant.

Given certain experimental manipulations, DCM models the activity of a set of ROIs at the neuronal level and then combines this hypothetical activity with a haemodynamic model. Then, it compares the modelled activity with the one obtained during the experiment and estimates coupling parameters between areas, so that the modelled activity is as similar as possible to the observed one. The details and the basic concepts of this procedure are explained below.

In DCM, F has the bilinear form:

$$\dot{z} = Az + \Sigma\, u_j\, B_j\, z + Cu$$

The parameters θ^n used to estimate the neuronal dynamics are:

1) Direct inputs on the state of any particular region (*C* matrix);
2) Intrinsic connections that couple the states between the regions (*A* matrix). These parameters estimate the impact that one neural system exerts over another in the absence of experimental perturbations.
3) Bilinear terms that model the changes induced by the experimental manipulations in regional activity (within a region) or the connections between regions (a specific pathway. I.e. the connection between V1 and V5) (*B* matrix).

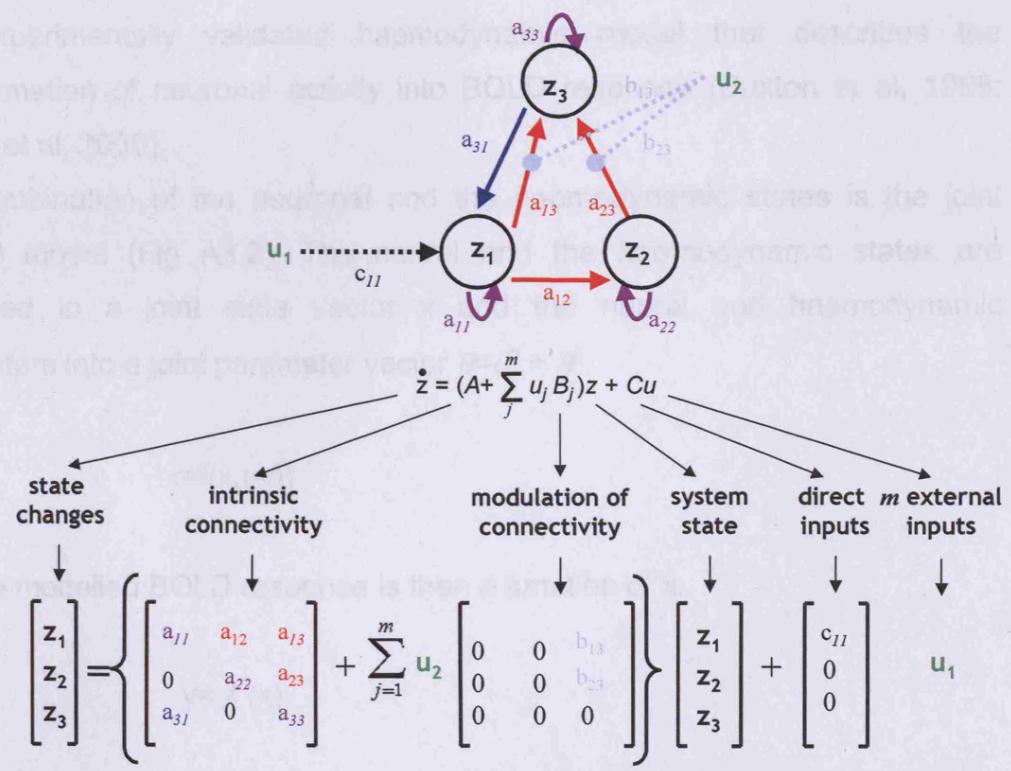


Fig A3.3. State equation. See text for details.

This can be visualized with the example in Fig A3.3. Here there are three areas of interest: z1, z2 and z3; for each one there is a times series of the BOLD responses (blue). In this model, z1 is connected via forward connections to z2 and z3 (represented by a12 and a13), and receives backward connections from z3 (a31). There is also a connection from z2 to z3 (a23).

U1 represents the external inputs into the system (i.e. the onsets of all visual stimulation), and affects only z1 directly. u2 is a modulatory effect (i.e. colour in the stimuli), that only modulates the forward connections from z1 and z2 to z3. These modulations (bilinear terms) are represented by the parameters b13 and b23.

The lower part of Fig A3.3 shows the neuronal states equation with each of these parameters.

Joint forward model

The neuronal dynamics model is then combined with a biophysically plausible and experimentally validated haemodynamic model that describes the transformation of neuronal activity into BOLD response (Buxton et al, 1998; Friston et al, 2000).

The combination of the neuronal and the haemodynamic states is the joint forward model (Fig A3.2). The neural and the haemodynamic states are combined in a joint state vector x and the neural and haemodynamic parameters into a joint parameter vector $\theta=\theta^n + \theta^h$.

Then

$$x=f(x,u,\theta)$$

And the modelled BOLD response is then a function of x.

$$y= \lambda (x)$$

A DCM is fitted to data by tuning the neurodynamic and haemodynamic parameters so as to minimise the discrepancy between predicted and observed fMRI time series. A Bayesian estimation scheme, through an expectation maximization (EM) algorithm, is used to determine the posterior density of the parameters, which is the probability distribution of a parameter in terms of its

mean and standard deviation. Therefore, for each parameter it is possible to compute the probability of a parameter being bigger than a certain threshold (i.e. the probability of a modulation being bigger than zero).

The parameters in DCM can be understood as rate constants of neural population responses that have an exponential nature. This is because the solution to a linear ordinary differential equation like (1) is an exponential function, where the parameters are inversely proportional to the half-life (τ) of the modelled neural responses. τ has units of seconds, therefore the units of the parameters are in Hertz (1/s) (Stephan et al, 2007).

Remarks

Since its introduction (Friston et al, 2003), DCM has been widely used in the analysis of neuroimaging data (For examples, see: Mechelli et al, 2003; Stephan et al, 2007). Its advantage is that it allows making inferences about causality, a feat that is not possible with standard methods of analysis of neuroimaging data or correlative measures of functional connectivity. DCM uses the temporal information contained in the fMRI time series to estimate parameters representing the amount of coupling between different brain areas and their modulation by experimental manipulations. This of particular relevance for the research presented in this thesis, since it become possible to make inferences about top-down and bottom-up mechanisms mediating the observed local changes in brain activity.

Part 5 – REFERENCES

Adams DL, and Zeki S. 2001. Functional organization of macaque V3 for stereoscopic depth. *Journal of neurophysiology* **86:** 2195-2203.

Adams RB, and Janata P. 2002. A comparison of neural circuits underlying auditory and visual object categorization. *NeuroImage* **16:** 361-377.

Adler A. 1944. Disintegration and restoration of optic recognition in visual agnosia: Analysis of a case. *Archives of Neurology and Psychiatry* **51:** 243–259.

Allison T, McCarthy G, Nobre A, Puce A, and Belger A. 1991. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral cortex* **4:** 544-554.

Altmann CF, Bulthoff HH, and Kourtzi Z. 2003. Perceptual organization of local elements into global shapes in the human visual cortex. *Curr Biol* **13:** 342-349.

Amedi A, Jacobson G, Hendler T, Malach R, and Zohary E. 2002. Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral cortex* **12:** 1202-1212.

Amedi A, Stern W, Camprodon J, Bermpohl F, Merabet L, Rotman S, Hemond C, Meijer P, and Pascual-Leone A. 2007. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat Neurosci* **10:** 687-689.

Armstrong KM, Fitzgerald JK, and Moore T. 2006. Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron.* **50:** 791-798.

Armstrong KM, and Moore T. 2007. Rapid enhancement of visual cortical response discriminability by microstimulation of the frontal eye field. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 9499-9504.

Attwell D, and Iadecola C. 2002. The neural basis of functional brain imaging signals. *Trends in neurosciences* **25:** 621-625.

Avidan G, Levy I, Hendler T, Zohary E, and Malach R. 2003. Spatial vs. object specific attention in high-order visual areas. *NeuroImage.* **19:** 308-318.

Bahnsen R. 1928. Eine Untersuchung uber Symmetrie und Symmetrie bei visuellen Wahrnehmungen. *Z. Psychol.* **108.**

Baillet S, Mosher JC, and Leahy RM. 2001. Electromagnetic Brain Mapping. *IEEE Signal Processing Magazine* **18:** 14-30.

Baker CI, Behrmann M, and Olson CR. 2002. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat Neurosci* **5:** 1210-1216.

Ban H, Yamamoto H, Fukunaga M, Nakagoshi A, Umeda M, Tanaka C, and Ejima A. 2006. Toward a common circle: interhemispheric contextual modulation in human early visual areas. *The Journal of neuroscience* **26:** 8804-8809.

Bar M. 2003. A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of cognitive neuroscience.* **15:** 600-609.

Bar M. 2004. Visual objects in context. *Nature reviews. Neuroscience.* **5:** 617-629.

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Schmidt AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, and Halgren E. 2006. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 449-454.

Bar M, Tootell RB, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, and Dale AM. 2001. Cortical mechanisms specific to explicit visual object recognition. *Neuron* **29:** 529-535.

Barone P, Batardiere A, Knoblauch K, and Kennedy H. 2000. Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *The Journal of neuroscience* **20:** 3263-3281.

Bartels A, and Zeki S. 2000. The architecture of the colour centre in the human visual brain: new results and a review. *Eur J Neurosci* **12:** 172-193.

Baylis G, Driver J, Baylis L, and Rafal R. 1994. Reading of letters and words in a patient with Balint's syndrome. *Neuropsychologia* **32:** 1273-1286.

Beck P, and Kaas J. 1998. Cortical connections of the dorsomedial visual area in new world owl monkeys (Aotus trivirgatus) and squirrel monkeys (Saimiri sciureus). *The Journal of Comparative Neurology* **400:** 18-34.

Behrmann M, and Avidan G. 2005. Congenital prosopagnosia: face-blind from birth. *Trends in cognitive sciences* **9:** 180-187.

Benson DF, and Greenberg JP. 1969. Visual form agnosia. A specific defect in visual discrimination. *Archives of neurology* **20:** 82-89.

Benson DF, and Greenberg JP. 1968. Visual form agnosia: a specific defect in visual recognition. *Transactions of the American Neurological Association* **93:** 189-191.

Berkeley G. 3632. An essay towards a new theory of vision. *Classics in the history of psychology, by Christopher D. Green.*

Biederman I. 1987. Recognition-by-components: a theory of human image understanding. *Psychological review* **94:** 115-147.

Booth MC, and Rolls ET. 1998. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex* **8:** 510-523.

Boynton GM, Engel SA, Glover GH, and Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of neuroscience .* **16:** 4207-4221.

Brett M, Jl A, R V, and Jb P. 2002. Region of interest analysis using an SPM toolbox. *NeuroImage* **16.**

Brewer A, Liu J, Wade A, and Wandell B. 2005. Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature neuroscience* **8:** 1102-1109.

Brincat SL, and Connor C. 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature neuroscience* **7:** 880-886.

Brincat SL, and Connor CE. 2006. Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron.* **49:** 17-24.

Broca PP. 1861. Loss of Speech, Chronic Softening and Partial Destruction of the Anterior Left Lobe of the Brain *Bulletin de la Société Anthropologique*. 235-238.

Brodmann K. 1905. Beitrage zur histologichen Lokalisation der Grosshirnrinde Dritte Mitteilung: Die Rindenfelder de nierderen. *Affen. J. Psychol. Neurol. Lpz.* **47:** 607-638.

Buchel C, Holmes AP, Rees G, and Friston KJ. 1998. Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *NeuroImage.* **8:** 140-148.

Bullier J. 2004. Communications between cortical areas of the visual system. In: Chalupa LM and Werner JS, eds. *The visual neuroscience*. 1 ed. Cambridge, Massachusetts.: MIT press. 522-540.

Burkhalter A, and Bernardo KL. 1989. Organization of corticocortical connections in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America* **86:** 1071-1075.

Buxton R, Wong E, and Frank L. 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic resonance in medicine.* **39:** 855-864.

Callaway EM. 2004. Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural networks : the official journal of the International Neural Network Society* **17:** 625-632.

Callaway EM. 1998. Local circuits in primary visual cortex of the macaque monkey. *Annual review of neuroscience* **21:** 47-74.

Chun MM, and Jiang Y. 1998. Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive psychology.* **36:** 28-71.

Cohen L, Martinaud O, Lemer C, Lehéricy S, Samson Y, Obadia M, Slachevsky A, and Dehaene S. 2003. Visual word recognition in the left and right hemispheres: anatomical and functional correlates of peripheral alexias. *Cerebral cortex* **13:** 1313-1333.

Connor CE, Brincat SL, and Pasupathy A. 2007. Transformation of shape information in the ventral pathway. *Current opinion in neurobiology* **17:** 140-147.

Cornsweet TN. 1970. *Visual Perception.* Academic Press., New York.

Daniel P, and Whitteridge D. 1961. The representation of the visula field on the cerebral cortex in monkeys. *J Physiol* **159:** 203-221.

de Lafuente V, and Romo R. 2005. Neuronal correlates of subjective sensory experience. *Nature neuroscience* **8:** 1698-1703.

de Lafuente V, and Romo R. 2006. Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 14266-14271.

Dejerine J. 1892. Contribution a l'etude anatomo-patologique et clinique des differentes varietes de cecite verbale. *Mem Soc Biol* **4:** 61-90.

Desimone R, Albright TD, Gross CG, and Bruce C. 1984. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* **4:** 2051-2062.

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, and Neitz J. 1996. Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America* **93:** 2382-2386.

Distler C, Boussaoud D, Desimone R, and Ungerleider LG. 1993. Cortical connections of inferior temporal area TEO in macaque monkeys. *The Journal of comparative neurology* **334:** 125-150.

Dobbins A, Zucker SW, and Cynader MS. 1987. Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature* **329:** 438-441.

Doniger GM, Foxe JJ, Murray MM, Higgins BA, Snodgrass JG, Schroeder CE, and Javitt DC. 2000. Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes. *Journal of cognitive neuroscience* **12:** 615-621.

Doniger GM, Foxe JJ, Schroeder CE, Murray MM, Higgins BA, and Javitt DC. 2001. Visual perceptual learning in human object recognition areas: a repetition priming study using high-density electrical mapping. *NeuroImage* **13:** 305-313.

Downing P, Jiang Y, Shuman M, and Kanwisher N. 2001. A cortical area selective for visual processing of the human body. *Science (New York, N.Y.)* **293:** 2470-2473.

Duchaine B, and Nakayama K. 2006. Developmental prosopagnosia: a window to content-specific face processing. *Current opinion in neurobiology* **16:** 166-173.

Dumoulin SO, and Hess RF. 2007. Cortical specialization for concentric shape processing. *Vision research* **47:** 1608-1613.

Edelman S, and Intrator N. 2000. (Coarse coding of shape fragments) + (retinotopy) approximately = representation of structure. *Spatial vision.* **13:** 255-264.

Efron R. 1968. What is perception? In: Cohen R and Wartofsky M, eds. *Boston studies in the philosophy of science*. New York: Humanities Press. 137–173.

Eger E, Henson R, Driver J, and Dolan R. 2007. Mechanisms of top-down facilitation in perception of visual objects studied by FMRI. *Cerebral cortex* **17:** 2123-2133.

Engel SA, Glover GH, and Wandell BA. 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex* **7:** 181-192.

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, and Shadlen MN. 1994. fMRI of human visual cortex. *Nature* **369:** 525.

Epstein R, and Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature* **392:** 598-601.

Evett LJ, and Humphreys GW. 1981. The use of abstract graphemic information in lexical access. *Quarterly Journal of Experimental Psychology* **33:** 325-350.

Falchier A, Clavagnier S, Barone P, and Kennedy H. 2002. Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of neuroscience* . **22:** 5749-5759.

Farah M. 1990. *Visual Agnosia*. MIT Press.

Farah M. 2004. *Visual Agnosia*. MIT Press.

Farah M, Levinson K, and Klein K. 1995. Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia* **33:** 661-674.

Felleman D, Burkhalter A, and van Essen D. 1997. Cortical connections of areas V3 and VP of macaque monkey extrastriate visual cortex. *The Journal of comparative neurology* **379:** 21-47.

Felleman D, and Van Essen D. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex* **1:** 1-47.

Flourens P. 1824. *Recherches expérimentales sur les propriétés et les fonctions du système nerveux, dans les animaux vertébrés.* J.B.Bailliere.

Foerster O. 1890. Ueber Rindenblindheit. *Albrecht v. Graefes Arch. Ophtal.* **36:** 94-108.

Forster KI, and Davis C. 1984. Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, & Cognition.* **10:** 680-698.

Foxe JJ, Murray MM, and Javitt DC. 2005. Filling-in in schizophrenia: a high-density electrical mapping and source-analysis investigation of illusory contour processing. *Cerebral cortex* **15:** 1914-1927.

Freedman DJ, and Assad JA. 2006. Experience-dependent representation of visual categories in parietal cortex. *Nature.* **443:** 85-88.

Freedman DJ, Riesenhuber M, Poggio T, and Miller EK. 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of neuroscience* **23:** 5235-5246.

Friston KJ. 2002. Functional integration and inference in the brain. *Progress in neurobiology.* **68:** 113-143.

Friston KJ. 2003. Learning and inference in the brain. *Neural Netw* **16:** 1325-1352.

Friston KJ. 2005. A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* **360:** 815-836.

Friston KJ, Harrison L, Daunizeau J, Kiebel S, Phillips C, Trujillo-Barreto N, Henson R, Flandin G, and Mattout Jrm. 2007. Multiple sparse priors for the M/EEG inverse problem. *Neuroimage.*

Friston KJ, Jezzard P, and Turner R. 1994. Analysis of functional MRI time-series. *Human Brain Mapping* **1:** 153-171.

Friston KJ, Penny W, and Glaser D. 2005. Conjunction revisited. *NeuroImage* **25:** 661-667.

Friston KJ, Harrison L, and Penny W. 2003. Dynamic causal modelling. *Neuroimage* **19:** 1273-1302.

Friston KJ, and Henson RN. 2006. Commentary on: Divide and conquer; a defence of functional localisers. *NeuroImage.* **30:** 1097-1099.

Friston KJ, Holmes AP, and Worsley KJ. 1999. How many subjects constitute a study? *Neuroimage* **10:** 1-5.

Friston KJ, Mechelli A, Turner R, and Price CJ. 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage* **12:** 466-477.

Friston KJ, Rotshtein P, Geng JJ, Sterzer P, and Henson RN. 2006. A critique of functional localisers. *NeuroImage.* **30:** 1077-1087.

Fritsch G, and Hitzig E. 1870. Uber die elektrische Erregbakeit des Grosshirns. *Arch. f. Anat. Physiol. u. wiss. Med.* **37:** 300-332.

Fulton JF. 1928. Observations upon the vascularity of the human occipital lobe during visual acitivity. *Brain* **L1:** 310-320.

Gabrieli JD, Brewer JB, Desmond JE, and Glover GH. 1997. Separate neural bases of two fundamental memory processes in the human medial temporal lobe. *Science (New York, N.Y.)* **276:** 264-266.

Gauthier II. 2000. What constrains the organization of the ventral temporal cortex? *Trends Cogn Sci* **4:** 1-2.

Gerlach C, Aaside CT, Humphreys GW, Gade A, Paulson OB, and Law I. 2002. Brain activity related to integrative processes in visual object recognition: bottom-up integration and the modulatory influence of stored knowledge. *Neuropsychologia* **40:** 1254-1267.

Gerlach C, Law I, Gade A, and Paulson OB. 1999. Perceptual differentiation and category effects in normal object recognition: a PET study. *Brain : a journal of neurology* **122 (Pt 11):** 2159-2170.

Gilaie-Dotan S, Ullman S, Kushnir T, and Malach R. 2002. Shape-selective stereo processing in human object-related visual areas. *Hum Brain Mapp* **15:** 67-79.

Gilbert CD, and Sigman M. 2007. Brain states: top-down influences in sensory processing. *Neuron* **54:** 677-696.

Gilchrist ID, Humphreys GW, and Riddoch MJ. 1996. Grouping and extinction: evidence for low-level modulation of visual selection. *Cognitive Nuropsychology* **13:** 1223-49.

Green DM, and Swets JA. 1966. *Signal Detection Theory and Psychophysics.* Wiley.

Gregory RL. 1997. Knowledge in perception and illusion. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **352:** 1121-1127.

Grili-Spector K, and Kanwisher N. 2005. Visual recognition: as soon as you know it is there, you know what it is. *Psychol Sci.* **16:** 152-160.

Grill-Spector K, Knouf N, and Kanwisher N. 2004. The fusiform face area subserves face perception, not generic within-category identification. *Nature neuroscience* **7:** 555-562.

Grill-Spector K, Kourtzi Z, and Kanwlsher N. 2001. The lateral occipital complex and its role in object recognition. *Vision Res* **41:** 1409-1422.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, and Malach R. 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* **24:** 187-203.

Grill-Spector K, Kushnir T, Edelman S, Itzchak Y, and Malach R. 1998a. Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* **21:** 191-202.

Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzchak Y, and Malach R. 1998b. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp* **6:** 316-328.

Grlll-Spector K, Kushnir T, Hendler T, and Malach R. 2000. The dynamics of object-selective activation correlate with recognition performance in humans. *Nature neuroscience.* **3:** 837-843.

Grlll-Spector K, and Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta psychologica* **107:** 293-321.

Gross CG, Bender DB, and Rocha-Mlranda CE. 1969. Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* **166:** 1303-1306.

Grossberg S. 1994. 3-D vision and figure-ground separation by visual cortex. *Perception & psychophysics* **55:** 48-121.

Grossberg S. 1999. The link between brain learning, attention, and consciousness. *Consciousness and cognition* **8:** 1-44.

Hagler JDJ, and Sereno MI. 2006. Spatial maps in frontal and prefrontal cortex. *NeuroImage* **29:** 567-577.

Han S, Jiang Y, Humphreys G, Zhou T, and Cai P. 2005a. Attentional modulation of perceptual grouping in human visual cortex: functional MRI studies. *Human Brain Mapping* **25:** 424-32.

Han S, Jiang Y, Mao L, Humphreys G, and Gu H. 2005b. Attentional modulation of perceptual grouping in human visual cortex: ERP studies. *Human Brain Mapping* **26:** 199-209.

Harrison LM, Stephan KE, Rees G, and Friston KJ. 2007. Extra-classical receptive field effects measured in striate cortex with fMRI. *NeuroImage* **34:** 1199-1208.

Hasson U, Hendler T, Ben Bashat D, and Malach R. 2001. Vase or face?: A neural correlate of shape-selective grouping processes in the human brain. *Journal of cognitive neuroscience* **13:** 744-753.

Haxby J, Grady C, Horwitz B, Ungerleider L, Mishkin M, Carson R, Herscovitch P, Schapiro M, and Rapoport S. 1991. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences of the United States of America* **88:** 1621-1625.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, and Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)* **293:** 2425-2430.

Heeger DJ. 1997. Signal detection theory. http://www.cns.nyu.edu/%7edavid/handouts/sdt-advanced.pdf

Heeger DJ, and Ress D. 2002. What does fMRI tell us about neuronal activity? *Nat Rev Neurosci* **3:** 142-151.

Hegde J, and Van Essen DC. 2000. Selectivity for complex shapes in primate visual area V2. *J Neurosci* **20:** RC61.

Henschen SE. 1893. On the visual path and centre. *Brain* **16.**

Henson RNA. 2003. Neuroimaging studies of priming. *Progress in neurobiology* **70:** 53-81.

Henson RNA, and Rugg MD. 2003. Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia* **41:** 263-270.

Holmes G. 1945. The Ferrier Lecture: The organization of the visual cortex in man. *Proc. Roy. Soc. Lond. B* **132:** 348-361.

Hopf J-M, Vogel E, Woodman G, Heinze H-J, and Luck SJ. 2002. Localizing visual discrimination processes in time and space. *Journal of neurophysiology* **88:** 2088-2095.

Horton JC. 1984. Cytochrome oxidase patches: a new cytoarchitectonic feature of monkey visual cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **304:** 199-253.

Howe CQ, and Purves D. 2005. Natural-scene geometry predicts the perception of angles and line orientation. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 1228-1233.

Hubel DH, and Wiesel TN. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* **160:** 106-154.

Hubel DH, and Wiesel TN. 1965. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J Neurophysiol* **28:** 229-289.

Hubel DH, and Wiesel TN. 1968. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology.* **195:** 215-243.

Hubel DH, and Wiesel TN. 1972. Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. *The Journal of comparative neurology* **146:** 421-450.

Hulme O. 2005. Visibility, Invisibility and Reportability. London: University of London.

Humphreys GW, Price CJ, and Riddoch MJ. 1999. From objects to names: a cognitive neuroscience approach. *Psychological research* **62:** 118-130.

Humphreys GW, and Riddoch MJ. 1987. *To see but not to see: a case study of visual agnosia.* Erlbaum, London.

Humphreys GW, Riddoch MJ, and Price CJ. 1997. Top-down processes in object identification: evidence from experimental psychology,

neuropsychology and functional anatomy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **352:** 1275-1282.

Hung CP, Kreiman G, Poggio T, and DiCarlo JJ. 2005. Fast readout of object identity from macaque inferior temporal cortex. *Science (New York, N.Y.)* **310:** 863-866.

Hupe JM, James AC, Payne BR, Lomber SG, Girard P, and Bullier J. 1998. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* **394:** 784-787.

Hutton C. 2000. 25-09-2000. SPM list. http://www.fil.ion.ucl.ac.uk/spm

Inouye T. 1909. *Die Sehstorungen bei Schussverletzungen der kortikalen Sehsphare nach Beobachtungen an Verwundeten der letzten japanischen Kriege.* W Engelmann, Leipzig.

Ishai A, Ungerleider LG, Martin A, Schouten JL, and Haxby JV. 1999. Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 9379-9384.

Ito M, and Komatsu H. 2004. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *J Neurosci* **24:** 3313-3324.

James TW, Humphrey GK, Gati JS, Servos P, Menon RS, and Goodale MA. 2002. Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia.* **40:** 1706-1714.

Janssen P, Vogels R, and Orban GA. 2000a. Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science (New York, N.Y.)* **288:** 2054-2056.

Janssen P, Vogels R, and Orban GA. 2000b. Three-dimensional shape coding in inferior temporal cortex. *Neuron* **27:** 385-397.

Johnson JS, and Olshausen BA. 2003. Timecourse of neural signatures of object recognition. *Journal of vision* **3:** 499-512.

Kanizsa G, and Gerbino W. 1976. Convexity and symmetry in figure-ground organization. In: Henle M, ed. *Vision and artifact.* New York: Springer.

Kanwisher N. 2000. Domain specificity in face perception. *Nature neuroscience.* **3:** 759-763.

Kanwisher N, McDermott J, and Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialised for face perception. *J Neurosci* **17:** 4302-4311.

Kastner S, De Weerd P, and Ungerleider LG. 2000. Texture segregation in the human visual cortex: A functional MRI study. *J Neurophysiol* **83:** 2453-2457.

Kaufman L, and Lu ZL. 2003. Basics of neuromagnetism and magnectic source imaging. In: Lu ZL and Kaufman L, eds. *Magnetic source imaging of the human brain*: Lawrence Erlbaum Associates.

Kersten D, Mamassian P, and Yuille A. 2004. Object perception as Bayesian inference. *Annual review of psychology* **55:** 271-304.

Kiebel SJ, Daunizeau J, Phillips C, and Friston KJ. 2008. Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG. *NeuroImage* **39:** 728-741.

Kleinschmidt A, Buchel C, Zeki S, and Frackowiak RS. 1998. Human brain activity during spontaneously reversing perception of ambiguous figures. *Proceedings. Biological sciences / The Royal Society.* **265:** 2427-2433.

Kleinschmidt A, and Cohen L. 2006. The neural bases of prosopagnosia and pure alexia: recent insights from functional neuroimaging. *Current opinion in neurology* **19:** 386-391.

Kochukhova O, and Gredeback G. 2006. Learning about occlusion: Initial assumptions. *Cognition.*

Kofka K. 1935. *Principles of Gestalt Psychology.* Harcourt & Brace, New York.

Kohler W. 1929. *Gestalt Psychology.* Liveright, New York.

Kourtzi Z, Erb M, Grodd W, and Bulthoff HH. 2003a. Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cerebral cortex* **13:** 911-920.

Kourtzi Z, and Kanwisher N. 2001. Representation of perceived object shape by the human lateral occipital complex. *Science.* **293:** 1506-1509.

Kourtzi Z, Tolias AS, Altmann CF, Augath M, and Logothetis NK. 2003b. Integration of local features into global shapes: monkey and human FMRI studies. *Neuron* **37:** 333-346.

Lamme VA. 1995. The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of neuroscience .* **15:** 1605-1615.

Lamme VA, and Roelfsema PR. 2000. The distinct modes of vision offered by feedforward and recurrent processing. *Trends in neurosciences.* **23:** 571-579.

Landis T, Cummings J, Benson D, and Palmer E. 1986. Loss of topographic familiarity. An environmental agnosia. *Archives of neurology* **43:** 132-136.

Larsson J, and Heeger D. 2006. Two retinotopic visual areas in human lateral occipital cortex. *The Journal of neuroscience* **26:** 13128-13142.

Lee T, and Mumford D. 2003. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, image science, and vision* **20:** 1434-1448.

Lee T, Mumford D, Romero R, and Lamme V. 1998. The role of the primary visual cortex in higher level vision. *Vision research* **38:** 2429-2454.

Lehmann C, Vannini P, Wahlund L-O, Almkvist O, and Dierks T. 2006. Increased sensitivity in mapping task demand in visuospatial processing using reaction-time-dependent hemodynamic response predictors in rapid event-related fMRI. *NeuroImage* **31:** 505-512.

Leventhal A, Wang Y, Schmolesky M, and Zhou Y. 1998. Neural correlates of boundary perception. *Visual neuroscience* **15:** 1107-1118.

Lissauer H. 1890. Ein fall von seelendblindheit nebst einem beitrage zur theorie derselben. *Arch Psychiatr Nervenkr* **21**.

Liu J, Harris A, and Kanwisher N. 2002. Stages of processing in face perception: an MEG study. *Nature neuroscience.* **5:** 910-916.

Livingstone M, and Hubel D. 1988. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science (New York, N.Y.)* **240:** 740-749.

Logothetis N, Pauls J, Augath M, Trinath T, and Oeltermann A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature.* **412:** 150-157.

Logothetis NK. 2000. Object recognition: holistic representations in the monkey brain. *Spatial vision* **13:** 165-178.

Logothetis NK. 2002. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philosophical*

transactions of the Royal Society of London. Series B, Biological sciences **357:** 1003-1037.

Logothetis NK, Pauls J, and Poggio T. 1995. Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* **5:** 552-563.

Logothetis NK, and Sheinberg DL. 1996. Visual object recognition. *Annu Rev Neurosci* **19:** 577-621.

Lund JS. 1988. Anatomical organization of macaque monkey striate visual cortex. *Annual review of neuroscience* **11:** 253-288.

Luria A. 1959. Disorders of "simultaneous perception" in a case of bilateral occipito-parietal brain injury. *Brain* **82:** 437-49.

Lutkenhoner B. 2003. Magnetoencephalography and its Achilles' heel. *J Physiolo* **97:** 641-658.

Macmillan N, and Creelman CD. 2005. *Detection Theory: A User's guide.* Erlbaum., Mahwah, NJ.

Malach R, Levy I, and Hasson U. 2002. The topography of high-order human object areas. *Trends Cogn Sci* **6:** 176-184.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, and Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* **92:** 8135-8139.

Marr D. 1982. *Vision.* Freeman.

Marr D, and Hildreth E. 1980. Theory of edge detection. *Proc. Roy. Soc. Lond. B* **207:** 187-217.

Marr D, and Nishihara H. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B.* **200:** 269-294.

Martínez A, Anllo-Vento L, Sereno MI, Frank LR, Buxton RB, Dubowitz DJ, Wong EC, Hinrichs H, Heinze HJ, and Hillyard SA. 1999. Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature neuroscience.* **2:** 364-369.

Mathiesen C, Caesar K, Akgören N, and Lauritzen M. 1998. Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. *The Journal of physiology* **512 (Pt 2):** 555-566.

Mattout J, Phillips C, Penny WD, Rugg MD, and Friston KJ. 2006. MEG source localization under multiple constraints: an extended Bayesian framework. *NeuroImage* **30:** 753-767.

Maunsell JH, and Van Essen DC. 1983. The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of neuroscience* . **3:** 2563-2586.

McCarthy G, Puce A, Belger A, and Allison T. 1999. Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral cortex* **9:** 431-444.

McKeeff TJ, and Tong F. 2007. The timing of perceptual decisions for ambiguous face stimuli in the human ventral visual cortex. *Cerebral cortex* **17:** 669-678.

McKeefry DJ, Watson JD, Frackowiak RS, Fong K, and Zeki S. 1997. The activity in human areas V1/V2, V3, and V5 during the perception of coherent and incoherent motion. *NeuroImage* **5:** 1-12.

McRobbie DW, Moore EA, Graves MJ, and Prince MR. 2002. *MRI from picture to proton*. Cambridge Univesity Press.

Mechelli A, Price CJ, and Friston KJ. 2001. Nonlinear coupling between evoked rCBF and BOLD signals: a simulation study of hemodynamic responses. *NeuroImage* **14:** 862-872.

Mechelli A, Price CJ, Noppeney U, and Friston KJ. 2003. A dynamic causal modeling study on category effects: bottom-up or top-down mediation? *J Cogn Neurosci* **15:** 925-934.

Mendez M, and Cherrier M. 2003. Agnosia for scenes in topographagnosia. *Neuropsychologia* **41:** 1387-1395.

Messinger A, Squire LR, Zola SM, and Albright TD. 2001. Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 12239-12244.

Metzger F. 1953. *Gesetze des Schens*. Waldemar Kramer, Frankfurt-am-Main.

Milner AD, and Goodale MA. 1993. Visual pathways to perception and action. *Progress in brain research* **95:** 317-337.

Mirabella G, Bertini G, Samengo I, Kilavik BE, Frilli D, Della Libera C, and Chelazzi L. 2007. Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. *Neuron* **54:** 303-318.

Morinaga S. 1941. Beobachtungen uber die Grundlagen und Wirkungen anschaulich gleichmassiger Breite. *Arch. ges. Psychol.* **108.**

Mosso A. 1881. *Uber den Kreislauf des Blutes in menschlichen Gehirn.* Verlag von Veit, Leipzig.

Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, and Malach R. 2005. Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. *Science* **309:** 951-954.

Munk H. 1881. *Uber die Funktionen der Grosshirinrinde. English translation 'On the Functions of the Cortex'. Charles Thomas: Springfield, IL, 1960, pp 97–117.* A Hirschwald, Berlin.

Murray M, Imber M, Javitt D, and Foxe J. 2006. Boundary Completion Is Automatic and Dissociable from Shape Discrimination. *Journal of Neuroscience* **26:** 12043-12054.

Murray SO, Kersten D, Olshausen BA, Schrater P, and Woods DL. 2002. Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci U S A* **99:** 15164-15169.

Navon D. 1977. Forest before trees: the precedence of global features in visual perception. *Cognitive Psychology* **9:** 353-383.

Nowak LG, Munk MH, Girard P, and Bullier J. 1995. Visual latencies in areas V1 and V2 of the macaque monkey. *Visual neuroscience* **12:** 371-384.

Ogawa S, Lee TM, Kay AR, and Tank DW. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America.* **87:** 9868-9872.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, and Ugurbil K. 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America.* **89:** 5951-5955.

Palmer S. 1992. Common region: a new principle of perceptual grouping. *Cognitive psychology* **24:** 436-447.

Palmer SE. 1999. *Vision science -- photons to phenomenology.* MIT Press., Cambridge, Massachusetts.

Palmer SE, and Rock I. 1994. Rethinking perceptual organization. *Psychollogical Bulletin and Review* **1**.

Pasupathy A, and Connor CE. 1999. Responses to contour features in macaque area V4. *J Neurophysiol* **82:** 2490-2502.

Pasupathy A, and Connor CE. 2001. Shape representation in area V4: position-specific tuning for boundary conformation. *Journal of neurophysiology* **86:** 2505-2519.

Pasupathy A, and Connor CE. 2002. Population coding of shape in area V4. *Nat Neurosci* **5:** 1332-1338.

Patterson K, and Kay J. 1982. Letter-by-letter reading: psychological description of a neurological syndrome. *Quaterly Journal of Experimental Psychology.* **34A:** 411-431.

Penfield W, and Rasmussen T. 1968. *The cerebral cortex of man: A clinical study of localization of function.* Hafner Pub. Co., New York.

Penny W, Holmes AP, and Friston K. 2003. Random effects analysis. In: Frackowiak R, Friston K, Fristh C, Dolan R, Price C, Zeki S and Ashburner J, eds. *Human Brain Function*: Academic Press.

Peppiatt C, Howarth C, Mobbs P, and Attwell D. 2006. Bidirectional control of CNS capillary diameter by pericytes. *Nature* **advanced online publication:** 700-704.

Pernet C, Franceries X, Basan S, Cassol E, Demonet J, and Celsis P. 2004. Anatomy and time course of discrimination and categorization processes in vision: an fMRI study. *NeuroImage* **22:** 1563-1577.

Perruchet P, and Pacton S. 2006. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in cognitive sciences* **10:** 233-238.

Peterson MA, and Skow-Grant E. 2003. Memory and learning in figure-ground perception. In: Ross B and Irwin D, eds. *Cognitive vision: Psychology of learning and motivation.* 1-34.

Poghosyan V, and Ioannides AA. 2007. Precise mapping of early visual responses in space and time. *NeuroImage* **35:** 759-770.

Poliak S. 1932. the main afferent fiber systems of the cerebral cortex in primates. 370.

Pollen DA. 1999. On the neural correlates of visual perception. *Cereb Cortex* **9:** 4-19.

Pomerantz J, Sager L, and Stoever R. 1977. Perception of wholes and of their component parts: some configural superiority effects. *Journal of experimental psychology. Human perception and performance* **3:** 422-435.

Press WA, Brewer AA, Dougherty RF, Wade AR, and Wandell BA. 2001. Visual areas and spatial summation in human visual cortex. *Vision research.* **41:** 1321-1332.

Price C, Moore C, Humphreys G, Frackowiak R, and Friston K. 1996. The neural regions sustaining object recognition and naming. *Proceedings. Biological sciences / The Royal Society* **263:** 1501-1507.

Purves D, Lotto RB, Williams SM, Nundy S, and Yang Z. 2001. Why we see things the way we do: evidence for a wholly empirical strategy of vision. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **356:** 285-297.

Quiroga RQ, Reddy L, Kreiman G, Koch C, and Fried I. 2005. Invariant visual representation by single neurons in the human brain. *Nature* **435:** 1102-1107.

Rakison DH. 2007. Fast tracking: infants learn rapidly about object trajectories. *Trends in cognitive sciences* **11:** 140-142.

Rao RP, and Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* **2:** 79-87.

Reber AS. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* **6:** 855-863.

Reddy L, and Kanwisher N. 2006. Coding of visual objects in the ventral stream. *Current opinion in neurobiology* **16:** 408-414.

Ress D, and Heeger DJ. 2003. Neuronal correlates of perception in early visual cortex. *Nature neuroscience.* **6:** 414-420.

Reynolds JH, and Chelazzi L. 2004. Attentional modulation of visual processing. *Annual review of neuroscience.* **27:** 611-647.

Riddoch G. 1917. Dissociation of visual perception due to occipital injuries, with especial reference to appreciation of movement. *Brain* **40:** 15-57.

Riddoch MJ, and Humphreys GW. 1987. A case of integrative visual agnosia. *Brain* **110 (Pt 6):** 1431-1462.

Riddoch MJ, and Humphreys GW. 2003. Visual agnosia. *Neurologic clinics* **21:** 501-520.

Riddoch MJ, Humphreys GW, Akhtar N, Allen H, Bracewell M, and Schofield AJ. 2008. A tale of two agnosias: distinctions between integrative and form agnosia. *Cognitive Neuropsychology (in press).*

Riesenhuber M, and Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nature neuroscience.* **2:** 1019-1025.

Ritter W, Simson R, and Vaughan JHG. 1983. Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology* **20:** 168-179.

Ritter W, Simson R, and Vaughan JHG. 1988. Effects of the amount of stimulus information processed on negative event-related potentials. *Electroencephalography and clinical neurophysiology* **69:** 244-258.

Ritter W, Simson R, Vaughan JHG, and Macht M. 1982. Manipulation of event-related potential manifestations of information processing stages. *Science (New York, N.Y.)* **218:** 909-911.

Rockland K, and Pandya D. 1979. Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain research* **179:** 3-20.

Rockland K, and Pandya D. 1981. Cortical connections of the occipital lobe in the rhesus monkey: interconnections between areas 17, 18, 19 and the superior temporal sulcus. *Brain research* **212:** 249-270.

Roy CS, and Sherrington CS. 3790. On the Regulation of the Blood-supply of the Brain. *The Journal of physiology* **11:** 85-158.17.

Rubin E. 1921. Visuell wahrgenommene Figuren (Gyldendals, Copenhagen). In: Yantis S, ed. *Visual Perception Essential Readings.* Philadelphia, 2001: Psychology Press.

Rubin N. 2001. Figure and ground in the brain. *Nat Neurosci* **4:** 857-858.

Saffran J, Aslin R, and Newport E. 1996. Statistical learning by 8-month-old infants. *Science (New York, N.Y.)* **274:** 1926-1928.

Sakagami M, Ki T, Lauwereyns J, Koizumi M, Kobayashi S, and Hikosaka O. 2001. A code for behavioral inhibition on the basis of color, but not motion, in ventrolateral prefrontal cortex of macaque monkey. *The Journal of neuroscience .* **21:** 4801-4808.

Sakagami M, and Tsutsui K. 1999. The hierarchical organization of decision making in the primate prefrontal cortex. *Neuroscience research* **34:** 79-89.

Sakai K, and Miyashita Y. 1991. Neural organization for the long-term memory of paired associates. *Nature* **354:** 152-155.

Salin PA, and Bullier J. 1995. Corticocortical connections in the visual system: structure and function. *Physiological reviews.* **75:** 107-154.

Schacter D, Dobbins i, and Schnyer D. 2004. Specificity of priming: a cognitive neuroscience perspective. *Nature reviews. Neuroscience* **5:** 853-862.

Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, and Leventhal AG. 1998. Signal timing across the macaque visual system. *Journal of neurophysiology* **79:** 3272-3278.

Self MW, and Zeki S. 2005. The integration of colour and motion by the human visual brain. *Cereb Cortex* **15:** 1270-1279.

Seltzer B, and Pandya DN. 1978. Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain research* **149:** 1-24.

Sereno M, and Huang R-S. 2006. A human parietal face area contains aligned head-centered visual and tactile maps. *Nature neuroscience* **9:** 1337-1343.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, and Tootell RB. 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268:** 889-893.

Sereno MI, McDonald CT, and Allman JM. 1994. Analysis of retinotopic maps in extrastriate cortex. *Cereb Cortex* **4:** 601-620.

Sereno MI, Pitzalis S, and Martinez A. 2001. Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science (New York, N.Y.)* **294:** 1350-1354.

Sergent J, Ohta S, and MacDonald B. 1992. Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain : a journal of neurology* **115 Pt 1:** 15-36.

Sharon E, Galun M, Sharon D, Basri R, and Brandt A. 2006. Hierarchy and adaptivity in segmenting visual scenes. *Nature* **442:** 810-813.

Shipp S, and Zeki S. 1995. Segregation and convergence of specialised pathways in macaque monkey visual cortex. *Journal of anatomy.* **187 (Pt 3):** 547-562.

Sigala N, and Logothetis N. 2002. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415:** 318-320.

Simons J, Koutstaal W, Prince S, Wagner A, and Schacter D. 2003. Neural mechanisms of visual object priming: evidence for perceptual and semantic distinctions in fusiform cortex. *NeuroImage* **19:** 613-626.

Smith AT, Singh KD, Williams AL, and Greenlee MW. 2001. Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral cortex* **11:** 1182-1190.

Sokoloff L. 1981. Relationships among local functional activity, energy metabolism, and blood flow in the central nervous system. *Federation proceedings* **40:** 2311-2316.

Somers DC, Nelson SB, and Sur M. 1995. An emergent model of orientation selectivity in cat visual cortical simple cells. *The Journal of neuroscience.* **15:** 5448-5465.

Spiridon M, and Kanwisher N. 2002. How distributed is visual category information in human occipito-temporal cortex?: An fMRI study. *Neuron.* **35:** 1157-1165.

Squatrito S, Trotter Y, and Poggio GF. 1990. Influences of uniform and textured backgrounds on the impulse activity of neurons in area V1 of the alert macaque. *Brain research.* **536:** 261-270.

Stark CE, and Squire LR. 2000. Functional magnetic resonance imaging (fMRI) activity in the hippocampal region during recognition memory. *The Journal of neuroscience* **20:** 7776-7781.

Stephan K. 2007. Dynamic Causal Modeling for fMRI. In: The FIL Methods Group, ed. *SPM5 Manual*: http://www.fil.ion.ucl.ac.uk/spm/doc/manual.pdf.

Stephan KE, Marshall JC, Penny WD, Friston KJ, and Fink GR. 2007. Interhemispheric integration of visual processing during task-driven lateralization. *The Journal of neuroscience* **27:** 3512-3522.

Sugiura M, Friston K, Willmes K, Shah N, Zilles K, and Fink G. 2006. Analysis of intersubject variability in activation: An application to the incidental episodic retrieval during recognition test. *Hum Brain Mapp.*

Super H, Spekreijse H, and Lamme VA. Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience.* **4:** 304-310.

Super H, van der Togt C, Spekreijse H, and Lamme VAF. 2003. Internal state of monkey primary visual cortex (V1) predicts figure-ground perception. *The Journal of neuroscience .* **23:** 3407-3414.

Suzuki W, Matsumoto K, and Tanaka K. 2006. Neuronal responses to object images in the macaque inferotemporal cortex at different stimulus discrimination levels. *The Journal of neuroscience .* **26:** 10524-10535.

Talairach J, and Tournoux P. 1988. *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - an Approach to Cerebral Imaging.* Thieme Medical Publishers, New York.

Talbot SA, and Marshall UH. 1941. Physiological studies on neural mechanisms of visula localization and discrimination. *Amer J Ophthal* **24:** 1255-1264.

Tanaka K. 1993. Neuronal mechanisms of object recognition. *Science (New York, N.Y.)* **262:** 685-688.

Tanaka K. 2000. Mechanisms of visual object recognition studied in monkeys. *Spat Vis* **13:** 147-163.

Tanaka K, Saito H, Fukada Y, and Moriya M. 1991. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of neurophysiology* **66:** 170-189.

Tanner JWP, and Swets JA. 1954. A decision-making theory of visual detection. *Psychol Rev*: 401-409.

Teo PC, Sapiro G, and Wandell BA. 1997. Creating connected representations of cortical gray matter for functional MRI visualization. *IEEE Trans Med Imaging* **16:** 852-863.

Thomsen K, Offenhauser N, and Lauritzen M. 2004. Principal neuron spiking: neither necessary nor sufficient for cerebral blood flow in rat cerebellum. *J Physiol* **560:** 181-189.

Thorpe S, Fize D, and Marlot C. 1996. Speed of processing in the human visual system. *Nature* **381:** 520-522.

Thurstone LL. 1927. Psychophysical Analysis. *The American journal of psychology* **100:** 587-609.

Tootell RB, Mendola JD, Hadjikhani NK, Ledden PJ, Liu AK, Reppas JB, Sereno MI, and Dale AM. 1997. Functional analysis of V3A and related areas in human visual cortex. *J Neurosci* **17:** 7060-7078.

Tsao DY, Vanduffel W, Sasaki Y, Fize D, Knutsen TA, Mandeville JB, Wald LL, Dale AM, Rosen BR, Van Essen DC, Livingstone MS, Orban GA, and Tootell RBH. 2003. Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. *Neuron* **39:** 555-568.

Tucker T, and Fitzpatrick D. 2006. Luminance-Evoked Inhibition in Primary Visual Cortex: A Transient Veto of Simultaneous and Ongoing Response. *Journal of Neuroscience* **26:** 13537-13547.

Tyler CW, Likova LT, Kontsevich LL, and Wade AR. 2006. The specificity of cortical region KO to depth structure. *NeuroImage.* **30:** 228-238.

Ullman S. 2006. Object recognition and segmentation by a fragment-based hierarchy. *Trends Cogn Sci.*

Uttal WR. 1988. *On seeing forms.* Lawrence Erlbaum Associates, Hillsdale, NJ.

van Essen D, and Zeki S. 1978. The topographic organization of rhesus monkey prestriate cortex. *The Journal of physiology* **277:** 193-226.

Van Essen DC, Newsome WT, Maunsell JH, and Bixby JL. 1986. The projections from striate cortex (V1) to areas V2 and V3 in the macaque monkey: asymmetries, areal boundaries, and patchy connections. *The Journal of comparative neurology* **244:** 451-480.

Vanrullen R, and Thorpe SJ. 2001. Is it a bird?: Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception.* **30:** 655-668.

Verrey L. 1888. Hemiachromatopsie droite absolue. *Archs Ophtalmol (Paris)* **8:** 289-301.

Versavel M, Orban GA, and Lagae L. 1990. Responses of visual cortical neurons to curved stimuli and chevrons. *Vision Res* **30:** 235-248.

von der Heydt R, Peterhans E, and Baumgartner G. 1984. Illusory contours and cortical neuron responses. *Science (New York, N.Y.)* **224:** 1260-1262.

von Helmholtz H. 1924. Helmholtz's Treatise on Physiological Optics. Translated from the third German edition. Wisconsin: Optical Society of America (USA).

Vuilleumier P, Henson RN, Driver J, and Dolan RJ. 2002. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature neuroscience* **5:** 491-499.

Wandell B, Brewer A, and Dougherty R. 2005. Visual field map clusters in human cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **360:** 693-707.

Wandell BA, Chial S, and Backus BT. 2000. Visualization and measurement of the cortical surface. *J Cogn Neurosci* **12:** 739-752.

Warnking J, Dojat M, Guerin-Dugue A, Delon-Martin C, Olympieff S, Richard N, Chehikian A, and Segebarth C. 2002. fMRI retinotopic mapping--step by step. *Neuroimage* **17:** 1665-1683.

Warrington E, and Shallice T. 1980. Word-form dyslexia. *Brain : a journal of neurology* **103:** 99-112.

Wertheimer M. 1923. Laws of Organization in Perceptual Forms. In: Ellis W, ed. *A source book of Gestalt psychology. (1938).* London: Routledge & Kegan Paul. 71-88.

Worsley KJ. 1994. Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Adv Appl Probability* **26:** 13-42.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, and Evans AC. 1996. A Unified Statistical Approach for Determining Significant Signals in Images of Cerebral Activation. *Human Brain Mapping* **4:** 58-73.

Yao H, Shi L, Han F, Gao H, and Dan Y. 2007. Rapid learning in cortical coding of visual scenes. *Nat Neurosci* **10:** 772-778.

Yukie M, and Iwai E. 1981. Direct projection from the dorsal lateral geniculate nucleus to the prestriate cortex in macaque monkeys. *The Journal of comparative neurology* **201:** 81-97.

Zeki S. 1973. Colour coding in rhesus monkey prestriate cortex. *Brain research* **53:** 422-427.

Zeki S. 1974. Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *The Journal of physiology* **236:** 549-573.

Zeki S. 1978a. Functional specialisation in the visual cortex of the rhesus monkey. *Nature.* **274:** 423-428.

Zeki S. 1978b. The third visual complex of rhesus monkey prestriate cortex. *The Journal of physiology.* **277:** 245-272.

Zeki S. 1978c. Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *The Journal of Physiology Online* **277:** 273-290.

Zeki S. 1980. A direct projection from area V1 to area V3A of rhesus monkey visual cortex. *Proceedings of the Royal Society of London. Series B.* **207:** 499-506.

Zeki S. 1983. The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. *Proceedings of the Royal Society of London. Series B.* **217:** 449-470.

Zeki S. 1993. *Vision of the brain.* **Blackwell Science.**

Zeki S, Perry RJ, and Bartels A. 2003. The processing of kinetic contours in the brain. *Cerebral cortex* **13:** 189-202.

Zeki S, and Shipp S. 1988. The functional logic of cortical connections. *Nature* **335:** 311-317.

Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, and Frackowiak RS. 1991. A direct demonstration of functional specialization in human visual cortex. *The Journal of neuroscience* **11:** 641-649.

Zhou H, Friedman H, and von der Heydt R. 2000. Coding of border ownership in monkey visual cortex. *The Journal of neuroscience* **20:** 6594-6611.

Zimmerman J, and Silver A. 1966. Macroscopic quantum interference effects through superconducting point contacts. *Physical Review* **141:** 367-375.